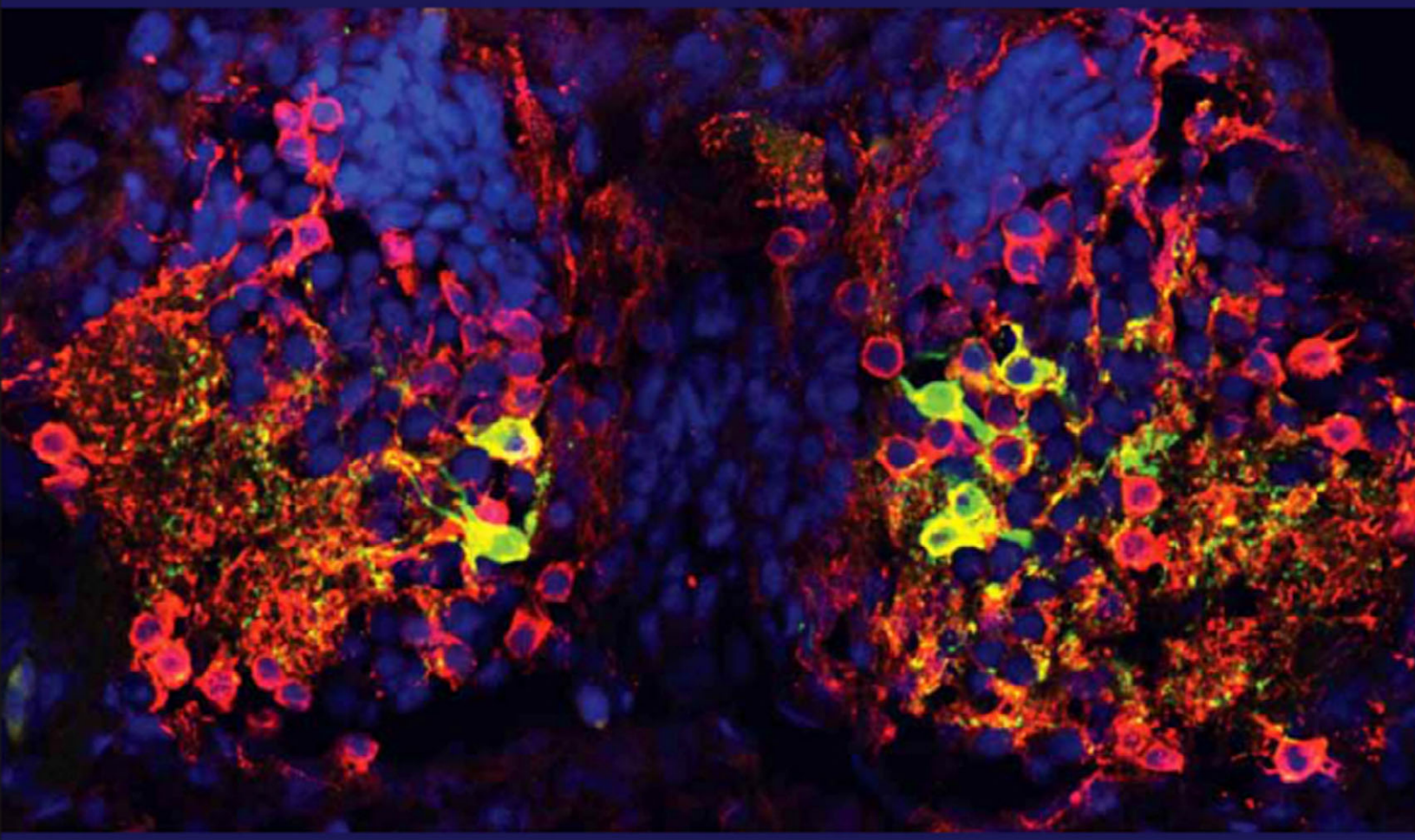
Fundamental Neuroscience



FOURTH EDITION

Larry R. Squire ■ Darwin Berg ■ Floyd E. Bloom
Sascha du Lac ■ Anirvan Ghosh
Nicholas C. Spitzer

AP

FUNDAMENTAL NEUROSCIENCE

Fourth Edition

This page intentionally left blank

FUNDAMENTAL NEUROSCIENCE

FOURTH EDITION

Edited by

LARRY R. SQUIRE
VA Medical Center San Diego, California
University of California, San Diego, La Jolla, California

DARWIN BERG
University of California, San Diego
La Jolla, California

FLOYD E. BLOOM
The Scripps Research Institute
La Jolla, California

SASCHA DU LAC
The Salk Institute
La Jolla, California

ANIRVAN GHOSH
University of California, San Diego
La Jolla, California

NICHOLAS C. SPITZER
University of California, San Diego
La Jolla, California

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Academic Press is an imprint of Elsevier

Academic Press is an imprint of Elsevier
225 Wyman Street, Waltham, MA 02451, USA
The Boulevard, Langford Lane, Kidlington, Oxford, OX5 1GB, UK

© 2013 Elsevier Inc. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the Publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: *www.elsevier.com/permissions*.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data
Fundamental neuroscience / edited by Larry Squire . . . [et al.]. – 4th ed.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-12-385870-2 (alk. paper)
1. Neurosciences. I. Squire, Larry R.
QP355.2.F862 2012
612.8–dc23 2012020285

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

Cover Image: The confocal image shows dopaminergic (green) and GABAergic (red) neurons in the developing Xenopus accessory olfactory bulb that mediate kinship recognition among siblings. The number of dopaminergic and GABAergic neurons changes in response to sustained exposure to olfactory stimuli. *Courtesy of Davide Dulcis and Nicholas Spitzer.*

For information on all Academic Press publications visit our website: *http://store.elsevier.com*

Printed in China
12 13 14 15 16 10 9 8 7 6 5 4 3 2 1

Working together to grow libraries in developing countries
www.elsevier.com | www.bookaid.org | www.sabre.org
ELSEVIER BOOK AID International Sabre Foundation

Short Contents

I
FUNDAMENTAL NEUROSCIENCE

1. Fundamentals of Neuroscience 3
2. Basic Plan of the Nervous System 15

II
CELLULAR AND MOLECULAR NEUROSCIENCE

3. Cellular Components of Nervous Tissue 41
4. Subcellular Organization of the Nervous System: Organelles and Their Functions 61
5. Membrane Potential and Action Potential 93
6. Neurotransmitters 117
7. Release of Neurotransmitters 139
8. Neurotransmitter Receptors 163
9. Intracellular Signaling 189
10. Postsynaptic Potentials and Synaptic Integration 211
11. Information Processing in Dendrites and Spines 231
12. Brain Energy Metabolism 261

III
NERVOUS SYSTEM DEVELOPMENT

13. Neural Induction and Pattern Formation 287
14. Cellular Determination 309
15. Neurogenesis and Migration 339
16. Growth Cones and Axon Pathfinding 363
17. Synapse Formation 385
18. Programmed Cell Death and Neurotrophic Factors 405
19. Synapse Elimination 437
20. Dendritic Development 457
21. Early Experience and Sensitive Periods 479

IV
SENSORY SYSTEMS

22. Fundamentals of Sensory Systems 499
23. Chemical Senses: Taste and Olfaction 513
24. The Somatosensory System 531
25. Audition 553
26. Vision 577

V
MOTOR SYSTEMS

27. Fundamentals of Motor Systems 599
28. The Spinal and Peripheral Motor System 613
29. Control of Movement 631
30. The Basal Ganglia 653
31. Cerebellum 677
32. Eye Movements 697

VI
REGULATORY SYSTEMS

33. The Hypothalamus: An Overview of Regulatory Systems 717
34. Central Control of Autonomic Functions: Organization of the Autonomic Nervous System 729
35. Neural Control of Respiratory and Cardiovascular Functions 749
36. Food Intake and Metabolism 767
37. Water and Salt Intake and Body Fluid Homeostasis 783
38. Neuroendocrine Systems 799
39. Circadian Timekeeping 819
40. The Neurobiology of Sleep and Dreaming 847
41. Reward, Motivation, and Addiction 871

VII
BEHAVIORAL AND COGNITIVE NEUROSCIENCE

42. Human Brain Evolution 901
43. Cognitive Development and Aging 919
44. Visual Perception of Objects 947
45. Spatial Cognition 969
46. Attention 989
47. Learning and Memory: Basic Mechanisms 1009
48. Learning and Memory: Brain Systems 1029
49. Language 1053
50. The Prefrontal Cortex and Executive Brain Functions 1069
51. The Neuroscience of Consciousness 1091

This page intentionally left blank

Full Contents

List of Boxes xv
Preface to the Fourth Edition xix
About the Editors xxi
Contributors xxiii

I
FUNDAMENTAL NEUROSCIENCE

1. Fundamentals of Neuroscience
FLOYD E. BLOOM

A Brief History of Neuroscience 3
The Terminology of Nervous Systems Is Hierarchical, Distributed, Descriptive, and Historically Based 3
Neurons and Glia Are Cellular Building Blocks of the Nervous System 4
Cellular Organization of the Brain 6
Organization of this Text 7
This Book Is Intended for a Broad Range of Scholars of the Neurosciences 8
Clinical Issues in the Neurosciences 8
The Genomic Inventory Is a Giant Step Forward 9
Neuroscience Today: A Communal Endeavor 10
The Creation of Knowledge 10
Responsible Conduct 11
Summary 12
References 12
Suggested Readings 12

2. Basic Plan of the Nervous System
LARRY W. SWANSON

Introduction 15
Evolution Highlights: General Organizing Principles 15
Development Reveals Basic Vertebrate Parts 23
The Basic Plan of Nervous System Connectivity 27
Overview of the Adult Mammalian Nervous System 31
References 37
Suggested Readings 38

II
CELLULAR AND MOLECULAR NEUROSCIENCE

3. Cellular Components of Nervous Tissue
PATRICK R. HOF, GRAHAME KIDD, JAVIER DEFELIPE, JEAN DE VELLIS, MIGUEL A. GAMA SOSA, GREGORY A. ELDER, AND BRUCE D. TRAPP

Neurons 41
Specific Examples of Different Neuronal Types 45
Neuroglia 49
Cerebral Vasculature 55
References 58
Suggested Readings 58

4. Subcellular Organization of the Nervous System: Organelles and Their Functions
SCOTT T. BRADY, DAVID R. COLMAN, AND PETER J. BROPHY

Axons and Dendrites: Unique Structural Components of Neurons 61
Protein Synthesis in Nervous Tissue 65
Cytoskeletons of Neurons and Glial Cells 75
Molecular Motors in the Nervous System 82
Building and Maintaining Cells of the Nervous System 85
References 91

5. Membrane Potential and Action Potential
DAVID A. MCCORMICK

Membrane Potential 94
Passive Electrical Properties of the Neuron 99
Action Potential 102
References 116
Suggested Readings 116

6. Neurotransmitters
ARIEL Y. DEUTCH

Several Modes of Neuronal Communication Exist 117
Chemical Transmission 118
Classical Neurotransmitters 120

Nonclassical Neurotransmitters 129
Peptide Transmitters 130
Unconventional Transmitters 132
The Expanding Synapse and Gliotransmission 136
References 138

7. Release of Neurotransmitters

THOMAS L. SCHWARZ

Transmitter Release Is Quantal 139
Excitation–Secretion Coupling 142
Molecular Mechanisms of the Nerve Terminal 145
Quantal Analysis: Probing Synaptic Physiology 155
Short-Term Synaptic Plasticity 158
References 161
Suggested Readings 161

8. Neurotransmitter Receptors

M. NEAL WAXHAM

Ionotropic Receptors 163
G-Protein Coupled Receptors 177
References 187

9. Intracellular Signaling

HOWARD SCHULMAN

Signaling Through G-Protein-Linked Receptors 189
Modulation of Neuronal Function by Protein Kinases and Phosphatases 197
Intracellular Signaling Affects Nuclear Gene Expression 205
References 208
Suggested Readings 209

10. Postsynaptic Potentials and Synaptic Integration

JOHN H. BYRNE

Ionotropic Receptors: Mediators of Fast Excitatory and Inhibitory Synaptic Potentials 211
Metabotropic Receptors: Mediators of Slow Synaptic Potentials 224
Integration of Synaptic Potentials 226
Synaptic Plasticity 227
References 229
Suggested Readings 230

11. Information Processing in Dendrites and Spines

NELSON SPRUSTON, MICHAEL HÄUSSER, AND GREG STUART

Introduction 231
Synaptic Integration in Passive Dendrites 234
Synaptic Integration in Active Dendrites 239
Structure and Function of Dendritic Spines 250
Computations Performed by Dendrites 254
Summary 258
References 258
Suggested Readings 259

12. Brain Energy Metabolism

IGOR ALLAMAN AND PIERRE J. MAGISTRETTI

Energy Metabolism of the Brain as a Whole Organ 261
Tight Coupling of Neuronal Activity, Blood Flow, and Energy Metabolism 264
Energy-Producing and Energy-Consuming Processes in the Brain 269
Glutamate Metabolism: A Coordinated Shuttle Between Astrocytes and Neurons 274
Brain Energy Metabolism at the Cellular Level 276
The Astrocyte-Neuron Metabolic Unit 283
References 284

III

NERVOUS SYSTEM DEVELOPMENT

13. Neural Induction and Pattern Formation

ANDREW LUMSDEN AND CLEMENS KIECKER

Neural Induction 287
Early Neural Patterning 292
Regionalization of the Central Nervous System 296
Conclusions 306
References 306

14. Cellular Determination

WILLIAM A. HARRIS, VOLKER HARTENSTEIN, AND MARTYN GOULDING

Introduction 309
Neurogenesis 310
The Proneural and Neurogenic Genes 312
Transcriptional Hierarchies and Networks 314
Asymmetric Cell Division and Cell Fate 314
Neurons and Glia 316
Sensory Neurons of the Peripheral Nervous System 320
The Retina 323
The Spinal Cord 328
Motor Neurons 330
The Cerebral Cortex 332
Conclusions 335
References 335

15. Neurogenesis and Migration

MARIANNE BRONNER AND MARY E. HATTEN

Introduction 339
Development of the Peripheral Nervous System 339
Segmental Migration of Neural Crest Cells 345
Cell Migration in the CNS 347
References 357
Suggested Readings 361

16. Growth Cones and Axon Pathfinding

ALEX L. KOLODKIN AND MARC TESSIER-LAVIGNE

Growth Cones Are Actively Guided 363
Guidance Cues for Developing Axons 366
Guidance Cues and the Control of Cytoskeletal Dynamics 374
Guidance at the Midline: Changing Responses to Multiple Cues 378
Topographic Mapping 380
References 382

17. Synapse Formation

STEVEN J. BURDEN AND PETER SCHEIFFELE

Development of the Neuromuscular Synapse 385
Synapse Formation in the Central Nervous System 394
References 402

18. Programmed Cell Death and Neurotrophic Factors

RONALD W. OPPENHEIM, CAROLANNE E. MILLIGAN, AND CHRISTOPHER S. VON BARTHELD

Cell Death and the Neurotrophic Hypothesis 406
The Origins of Programmed Cell Death and Its Widespread Occurrence in the Developing Nervous System 407
Functions of Neuronal Programmed Cell Death 409
Modes of Cell Death in Developing Neurons 410
The Mode of Neuronal Cell Death Reflects the Activation of Distinct Biochemical and Molecular Mechanisms 412
Nerve Growth Factor: The Prototype Target-Derived Neuronal Survival Factor 416
The Neurotrophin Family 418
Neurotrophin Receptors 419
Secretion and Axonal Transport of Neurotrophins and Pro-Neurotrophins 421
Signal Transduction Through TRK Receptors 423
Multiple Cytokines and Growth Factors Implement Neurotrophic Activities 425
Programmed Cell Death Is Regulated by Interactions with Targets, Afferents, and Nonneuronal Cells 429
The Role of Trophic Factors and Programmed Cell Death in Neuropathology 431
References 434
Suggested Readings 435

19. Synapse Elimination

JUAN C. TAPIA AND JEFF W. LICHTMAN

Overview 437
The Purpose of Synapse Elimination 441
A Structural Analysis of Synapse Elimination at the Neuromuscular Junction 443
A Role for Interaxonal Competition and Activity 446
Is Synapse Elimination Strictly a Developmental Phenomenon? 453
References 454

20. Dendritic Development

HOLLIS CLINE, ANIRVAN GHOSH, YUH-NUNG JAN, AND JOSH SANES

Dynamics of Dendritic Arbor Development 457
Transcriptional Control of Dendrite Development 458
Extracellular Regulation of Dendritic Development 460
Dendritic Development and Circuit Formation in Mammalian Retina 466
Laminar Specificity in the Inner Plexiform Layer 466
Activity-Dependent Dendritic Development 469
The Synaptotrophic Model of Dendrite Development 469
Branch Retraction and Synapse Elimination 471
Calcium-Dependent Mechanisms that Mediate Dendritic Growth 471
Protein Synthesis Dependent Regulation of Dendrite Development 472
Convergence and Divergence 472
Conclusion 473
References 474

21. Early Experience and Sensitive Periods

ERIC I. KNUDSEN

Birdsong: Learned by Experience 481
Sound Localization: Calibrated by Early Experience in the Owl 483
Binocular Processing in the Mammalian Visual Cortex Depends on Early Experience 487
A Sensitive Period for Shaping the Temperament of Rats 489
Principles of Sensitive Period Learning 491
References 494
Suggested Readings 495

IV
SENSORY SYSTEMS

22. Fundamentals of Sensory Systems

STEWART H. HENDRY AND STEVEN S. HSIAO

Sensation and Perception 499
Receptors 501
Peripheral Organization and Processing 503
Central Pathways and Processing 505
Sensory Cortex 507
Summary 510
References 510
Suggested Readings 511

23. Chemical Senses: Taste and Olfaction

KRISTIN SCOTT

Taste 513
Olfaction 520

Different Odors Activate Different Combinations of Glomeruli 524
Pheromone Detection 527
References 529
Suggested Readings 530

24. The Somatosensory System

STEWART HENDRY AND STEVEN HSIAO

Peripheral Mechanisms of Somatic Sensation 531
Nociception 538
CNS Components of Somatic Sensation 540
Thalamic Mechanisms of Somatic Sensation 544
The Path from Nociception to Pain 545
Cortical Representation of Touch 549
References 551
Suggested Readings 551

25. Audition

M. CHRISTIAN BROWN AND JOSEPH SANTOS-SACCHI

External and Middle Ear 553
The Cochlea 554
The Auditory Nerve 562
Central Nervous System 566
References 575
Suggested Readings 576

26. Vision

R. CLAY REID AND W. MARTIN USREY

Overview 577
Sensory Systems Detect Contrast or Change 578
The Eye and the Retina 579
The Retinogeniculocortical Pathway 586
References 595
Suggested Readings 595

V
MOTOR SYSTEMS

27. Fundamentals of Motor Systems

STEN GRILLNER

Basic Components of the Motor System 600
Motor Programs Coordinate Basic Motor Patterns 602
Roles of Different Parts of the Nervous System in the Control of Movement 605
Conclusion 611
Suggested Readings 611

28. The Spinal and Peripheral Motor System

GEORGE Z. MENTIS

Locomotion Is a Cycle 613
Connecting the Spinal Cord to the Periphery 615
Spinal Interneuron Networks 622
Descending Control of Spinal Circuits 627
Sensory Modulation 628
References 629
Suggested Readings 630

29. Control of Movement

MARC H. SCHIEBER AND JAMES F. BAKER

The Medial Postural System 631
The Lateral Voluntary System 638
Summary 650
References 650

30. The Basal Ganglia

JONATHAN W. MINK

Anatomy of the Basal Ganglia 653
Signaling in Basal Ganglia 661
The Effect of Basal Ganglia Damage on Behavior 664
Principles of Basal Ganglia Operation for Motor Control 668
Basal Ganglia Participation in Nonmotor Functions 669
References 675
Suggested Readings 676

31. Cerebellum

MARTHA BAGNALL, SASCHA DU LAC, AND MICHAEL MAUK

Cerebellar Anatomy and Circuit 677
Cellular and Synaptic Performance of the Cerebellar Circuit 684
Cerebellar Function 687
References 695

32. Eye Movements

RICHARD J. KRAUZLIS

Eye Movements Are Used to Stabilize Gaze or to Redirect Gaze 697
The Mechanics of Moving the Eyes 700
The Fundamental Circuits for Stabilizing Gaze 703
The Commands for Redirecting Gaze are Formed in the Brainstem 705
Gaze Movements are Controlled by the Midbrain and Forebrain 707

The Control of Gaze Movements Involves Higher-Order Processes 711
Conclusions 713
References 713
Suggested Readings 714

VI
REGULATORY SYSTEMS

33. The Hypothalamus: An Overview of Regulatory Systems

J. PATRICK CARD AND LARRY W. SWANSON

Historical Perspective 717
Hypothalamic Cytoarchitecture 718
Functional Organization of the Hypothalamus 719
Effector Systems of the Hypothalamus are Hormonal and Synaptic 722
References 726
Suggested Readings 727

34. Central Control of Autonomic Functions: Organization of the Autonomic Nervous System

TERRY L. POWLEY

Sympathetic Division: Organized to Mobilize the Body for Activity 730
Parasympathetic Division: Organized for Energy Conservation 734
The Enteric Division of the ANS: The Nerve Net Found in the Walls of Visceral Organs 736
ANS Pharmacology: Transmitter and Receptor Coding 738
Autonomic Coordination of Homeostasis 740
Hierarchically Organized ANS Circuits in the CNS 743
Perspective: Future of the Autonomic Nervous System 746
References 747
Suggested Readings 747

35. Neural Control of Respiratory and Cardiovascular Functions

J.L. FELDMAN, D.R. MCCRIMMON, AND S.F. MORRISON

Breathing 750
Where Are the Neurons Generating the Respiratory Pattern? 751
Discharge Patterns of Respiratory Neurons 752
Models for Respiratory Pattern Generation 752
Mechanoreceptors in the Lungs Adjust Breathing Pattern and Initiate Protective Reflexes 756
Cardiovascular 757
Sensory Regulation of Respiratory and Cardiovascular Systems 761
Modulation and Plasticity of Respiratory Motor Output 762
References 765
Suggested Readings 766

36. Food Intake and Metabolism

STEPHEN C. WOODS AND EDWARD M. STRICKER

Caloric Homeostasis 767
Role of Caloric Homeostasis in Control of Food Intake 769
Central Control of Food Intake 776
Neuropeptides and the Control of Food Intake 777
References 781
Suggested Readings 781

37. Water and Salt Intake and Body Fluid Homeostasis

EDWARD M. STRICKER AND JOSEPH G. VERBALIS

Body Fluid Physiology 783
Osmotic Homeostasis 784
Volume Homeostasis 790
References 796
Suggested Readings 797

38. Neuroendocrine Systems

ANDREA C. GORE

The Hypothalamus is the Central Neuroendocrine Organ 799
Hypothalamic Control of Anterior Pituitary Hormones and Their Regulated Functions 800
Hypothalamic Control of Posterior Pituitary Hormones and Their Regulated Functions 813
Hormones, the Brain, and Behavior 815
References 817

39. Circadian Timekeeping

DAVID R. WEAVER AND PATRICK EMERY

Introduction 819
Overview of the Mammalian Circadian Timing System 819
A Hierarchy of Cell-Autonomous Circadian Oscillators 821
The Molecular Basis for Circadian Oscillation Is a Transcriptional Feedback Loop 824
Circadian Photoreception in Mammals 832
Circadian Output Mechanisms 836
Diversity of Output Pathways Leading to Physiological Rhythms 837
General Summary 843
References 844
Suggested Readings 845

40. The Neurobiology of Sleep and Dreaming

EDWARD F. PACE-SCHOTT AND J. ALLAN HOBSON

The Two States of Sleep: Rapid Eye Movement (REM) and Non–Rapid Eye Movement (NREM) Sleep 847
Neural Control of the Sleep–Wake and REM/NREM Cycles 850
Systems Neuroscience of Sleep and Dreaming 859

Development and Phylogeny of Sleep 861
Sleep Disorders 862
The Purpose of Sleep 865
References 867
Suggested Readings 869

41. Reward, Motivation, and Addiction
GEORGE F. KOOB, BARRY J. EVERITT, AND TREVOR W. ROBBINS

Reward and Motivation 871
Addiction 879
References 896

VII
BEHAVIORAL AND COGNITIVE NEUROSCIENCE

42. Human Brain Evolution
JON H. KAAS AND TODD M. PREUSS

Human Brain Evolution 901
The Brains of Early Mammals 901
Evolution of Primate Brains 906
Conclusions 916
References 917
Suggested Readings 918

43. Cognitive Development and Aging
PETER R. RAPP AND JOCELYNE BACHEVALIER

Brain Development 919
Cognitive Development and Aging: A Life Span Perspective 923
Cognitive Aging 927
Pathological Processes in Cognitive Development and Aging 934
Pathological Manifestations of Cognitive Aging: Dementia 940
References 943
Suggested Readings 944

44. Visual Perception of Objects
ANDREW H. BELL, LUIZ PESSOA, ROGER B.H. TOOTELL, AND LESLIE G. UNGERLEIDER

The Problem of Object Recognition 947
Substrates for Object Perception and Recognition: Early Evidence from Brain Damage 948
Visual Pathways for Object Processing in Nonhuman Primates 950
Neuronal Properties within the Object Recognition Pathway 953
Functional Neuroimaging of Object Recognition 959
Perception and Recognition of Specific Classes of Objects 960
Overall Summary 966
References 966
Suggested Readings 968

45. Spatial Cognition
CARL R. OLSON AND CAROL L. COLBY

Neural Systems for Spatial Cognition 969
Parietal Cortex 970
Frontal Cortex 978
Medial Temporal Lobe 984
Spatial Cognition and Spatial Action 986
References 986
Suggested Readings 987

46. Attention
JOHN H. REYNOLDS, JACQUELINE P. GOTTLIEB, AND SABINE KASTNER

Introduction 989
Definitions and Varieties of Attention 989
Neglect Syndrome: A Deficit of Spatial Attention 990
Attention Affects Neural Activity in the Human Visual Cortex in the Presence and Absence of Visual Stimulation 992
Single Unit Recording Studies in Nonhuman Primates Provide Convergent Evidence for a Fronto-Parietal Attentional Control System 994
Attention Increases Neuronal Responses and Boosts the Clarity of Signals Generated by Neurons in Parts of the Visual System Devoted to Processing Information about Objects 998
Attention Modulates Neural Responses in the Lateral Geniculate Nucleus 1000
The Visual Search Paradigm has been Used to Study the Role of Attention in Selecting Relevant Stimuli from within a Cluttered Visual Environment 1001
Where Is the Computational Bottleneck as Revealed by Search Tasks? 1001
Neuronal Receptive Fields Are a Possible Neural Correlate of Limited Capacity 1002
Filtering of Unwanted Information in Humans 1004
Competition Can Be Biased by Nonspatial Feedback 1005
Conclusions 1006
References 1006
Suggested Readings 1007

47. Learning and Memory: Basic Mechanisms
JOHN H. BYRNE

Paradigms Have Been Developed to Study Associative and Nonassociative Learning 1009
Invertebrate Studies: Key Insights from *Aplysia* into Basic Mechanisms of Learning 1010
Vertebrate Studies: Long-Term Potentiation 1016
Long-Term Depression 1023
How Does a Change in Synaptic Strength Store a Complex Memory? 1023
References 1026
Suggested Readings 1027

48. Learning and Memory: Brain Systems
JOSEPH R. MANNS AND ELIZABETH A. BUFFALO

Introduction 1029
Behavior Supported by Multiple Memory Systems 1048
Conclusion 1049
References 1049
Suggested Readings 1051

49. Language
DAVID N. CAPLAN AND JAMES L. GOULD

Animal Communication 1053
Human Language 1057
Neural Organization for Language 1060
Conclusions 1066
References 1067
Suggested Readings 1067

50. The Prefrontal Cortex and Executive Brain Functions
EARL K. MILLER AND JONATHAN D. WALLIS

Introduction 1069
Anatomy and Organization of the Prefrontal Cortex 1069
Effects of Damage to the Prefrontal Cortex in Humans 1071
Neuroimaging Studies and PFC 1074
Effects of Damage to the Prefrontal Cortex in Monkeys 1075
Neurophysiology of the Prefrontal Cortex 1078
Theories of Prefrontal Cortex Function 1083
References 1088
Suggested Readings 1089

51. The Neuroscience of Consciousness
CHRISTOF KOCH

What Phenomena Does Consciousness Encompass? 1091
The Neurobiology of Free Will 1092
Consciousness in Other Species 1093
Arousal and States of Consciousness 1093
The Neuronal Correlates of Consciousness 1096
The Neuronal Basis of Perceptual Illusions 1097
Perceptual Puzzles of Contemporary Interest 1098
Forward versus Feedback Projections 1099
An Information-Theoretical Theory of Consciousness 1100
Conclusion 1101
References 1102

Index 1105

This page intentionally left blank

List of Boxes

2.1: The Neuron Doctrine 16
2.2: Cajal: Iconoclast to Icon 20
2.3: Anatomical Relationships in the Vertebrate Body 21
5.1: Nernst Equation 96
5.2: Goldman-Hodgkin-Katz Equation 98
5.3: Voltage-Clamp Technique 103
5.4: Ion Channels and Disease 107
6.1: MAO and COMT Inhibitors in the Treatment of Neuropsychiatric Disorders 125
6.2: Going for Gases as Neurotransmitter 134
10.1: Optogenetics 213
11.1: Neuronal Cable Theory and Computational Modeling 232
11.2: Dendritic Recording 235
11.3: Imaging Dendritic Function 242
12.1: Positron Emission Tomography (PET) and Functional Magnetic Resonance Imaging (fMRI) 266
12.2: Hepatic Encephalopathy 275
14.1: Neural Stem Cells 318
15.1: Adult Neurogenesis 352
17.1: *Torpedo* Electric Organ 386
17.2: Synaptic Targets in Neurodevelopmental Disorders 398
19.1: α–Bungarotoxin 448
20.1: Rho GTPases Control the Structure of the Dendritic Cytoskeleton 464
21.1: Sensitive Periods in Humans 480
21.2: Filial Imprinting: Babies Learn to Recognize Their Parents 492
23.1: Identifying a Genetic Locus Linked to Bitter Taste Perception 516
23.2: Taste Cells in Fruit Flies 518
23.3: Sensing Thousands of Chemicals with a Handful of Neurons 522
25.1: Otoacoustic Emissions 560
25.2: Cochlear Implants 561
25.3: Giant Synaptic Terminals: Endbulbs and Calyces 568
25.4: Bat Echolocation 573
26.1: Kuffler's Study of Center-Surround Retinal Ganglion Cells 578
26.2: Quantitative Methods in the Study of Visual Neurons Classification of Retinal Ganglion Cells 582
26.3: Inherited and Acquired Defects of Color Vision: Retinal and Cortical Mechanisms 584
26.4: Vision in Mice 585
27.1: Central Pattern Generator Networks 604
28.1: Species Diversity in Neuromuscular Systems 614
28.2: Myasthenia and Myasthenic Disorders 618
28.3: Motor Neuron Diseases 622
28.4: Nonlinear Properties of Motor Neurons and CPG Interneurons 625
28.5: Neural Control of Human Walking 627
29.1: Vestibular Plasticity 637
29.2: Meniere Syndrome 638
29.3: Pyramids in the Brain 640

29.4: Motor Neuroprosthetics 648
30.1: Parkinson's Disease 657
30.2: The MPTP Story 667
30.3: Ventral System of the Basal Ganglia 670
30.4: Huntington's Disease 671
30.5: Deep Brain Stimulation 671
30.6: Obsessive-Compulsive Disorder 672
30.7: Tourette Syndrome 673
30.8: The Role of the Subthalamic Nucleus in Stopping 674
31.1: Parasagittal Stripes: The Basis of Cerebellar Microcircuits? 680
31.2: Differences Between the Vestibular and NonVestibular Portions of the Cerebellum 682
31.3: Clinical Testing Can Reveal Cerebellar Damage 688
31.4: Cognition, Emotion, and the Cerebellum 690
32.1: Eye Movements and Reading 700
32.2: Eye Movement Disorders 708
34.1: Autonomic Postganglionic Neurons Can Change Their Transmitter Phenotypes 739
34.2: Autonomic Reflexes: Activated by Visceral Afferents, Autonomic Afferents, or Just Afferents? 740
34.3: Diseases and Aging Impair the Autonomic Wisdom of the Body 744
34.4: Central ANS Circuits Integrate Autonomic Reflexes with Affective or Emotional Responses 745
35.1: Experimental Analysis of Ventilation 751
35.2: CNS Lesions Produce Abnormal Breathing Patterns 753
35.3: Why Two Distinct Oscillators? 755
35.4: Cardiovascular Disease 760
35.5: Disorders of Breathing Are Widespread and Serious 763
36.1: Validating Satiation Factors 771
36.2: The Glucostatic Hypothesis 773
36.3: Treatment of Obesity 774
37.1: Water Channels 786
37.2: Diabetes Insipidus 788
37.3: Water Intoxication 791
37.4: Salt Appetite After Adrenocortical Dysfunction 794
37.5: Thirst and Salt Appetite During Hypovolemia in Rats 796
38.1: Historical Perspective on Hypothalamic Hormones 800
38.2: Endocrine-Disrupting Chemicals and Their Effects on Neuroendocrine Systems 816
39.1: Neuronal Control of *Drosophila* Circadian Behavior 823
39.2: Molecular Mechanisms of Circadian Timekeeping in *Drosophila* 826
39.3: Circadian-Based Sleep Disorders, Clock Genes and Neuropsychiatric Diseases 829
39.4: Chronopharmacology 838
39.5: Jet Lag and Shift Work 840
39.6: Lost Without a Clock 842
41.1: Logic of Methodology Used in Behavioral Neuroscience 874
41.2: Separation of Sensory and Response Factors in the Dopaminergic Syndrome of Neglect 876
41.3: Psychophysics of Hedonism as Measured by Brain Stimulation Reward 878
41.4: Impulsivity and Compulsivity 883
41.5: OPIOID Receptors 885
41.6: Allostasis 891
41.7: Craving 891
41.8: Major Depressive Disorders 894
42.1: How Do We Learn About Brain Evolution? 902
42.2: Why Brain Size Is Important 910
43.1: The Optical Fractionator Stereological Method 931
43.2: Dyslexia 935
43.3: Triplet Repeat Disorders 936

44.1: Case Study of Associative Agnosia 949
44.2: The Issue of Viewpoint Dependence in Object Recognition 955
44.3: Linking Category-Selective Regions into Specialized Networks 963
45.1: Area 7a Neurons Carry Object-Centered Signals in a Constructional Task 978
47.1: Is Memory More than Changes in Synaptic Strength? 1020
48.1: Alzheimer's Disease 1032
50.1: Functional Topography 1076
50.2: Attention Deficit Hyperactivity Disorder 1080

This page intentionally left blank

Preface to the Fourth Edition

In this fourth edition of *Fundamental Neuroscience*, we have tried to improve on the third edition with a volume that effectively introduces students to the full range of contemporary neuroscience. Neuroscience is a large field founded on the premise that all of behavior and all of mental life have their origin in the structure and function of the nervous system. Today, the need for a single-volume introduction to neuroscience is greater than ever. Toward the end of the twentieth century, the study of the brain moved from a peripheral position within both the biological and psychological sciences to become an interdisciplinary field that is now central within each discipline. The maturation of neuroscience has meant that individuals from diverse backgrounds—including molecular biologists, computer scientists, and psychologists—are interested in learning about the structure and function of the brain and about how the brain works. In addition, new techniques and tools have become available to study the brain in increasing detail. New genetic methods have been introduced to delete or overexpress single genes with spatial and temporal specificity. Neuroimaging techniques such as functional magnetic resonance imaging (fMRI) have been developed that allow study of the living human brain while it is engaged in cognition.

This fourth edition attempts to capture the promise and excitement of this fast-moving discipline. All the chapters have been rewritten and updated. The volume begins with an opening chapter that provides an overview of the discipline. A second chapter presents fundamental information about the architecture and anatomy of nervous systems. The remainder of the volume (Sections II–VII) presents the major topics of neuroscience. The second section (Cellular and Molecular Neuroscience) considers the cellular and subcellular organization of neurons, the physiology of nerve cells, and how signaling occurs between neurons. The third section (Nervous System Development) includes discussion of neural induction, cell fate, migration, pathfinding, synapse formation, programmed cell death, synapse elimination, development of dendrites, and early experience, including critical periods. The fourth and fifth sections (Sensory Systems and Motor Systems) describe the neural organization of each sensory modality and the organization of the brain pathways and systems important for locomotion, voluntary action, and eye movements. The sixth section (Regulatory Systems) describes the variety of hypothalamic and extra-hypothalamic systems that support motivation, reward, and internal regulation, including respiration, food and water intake, neuroendocrine function, circadian rhythms, and sleep and dreaming. The final section (Behavioral and Cognitive Neuroscience) describes the neural foundations of the so-called higher mental functions, including perception, attention, memory, language, spatial cognition, and executive function. Additional chapters cover human brain evolution, cognitive development and aging, and consciousness. The chapters contain more than 100 Boxes that describe clinical conditions, techniques, and special topics related to chapter content. The volume is also accompanied by companion websites for students and instructors, which present all the Figures and Boxes. In addition, the instructor's website includes sample test questions for each chapter that draw attention to the key facts and concepts in the chapters.

The authors of the chapters and boxes are working scientists, experts in the topics they cover. The Editors edited the chapters to achieve consistency of style and content. At Academic Press/Elsevier Science, the project was coordinated by Mica Haley (Acquisitions Editor), and we are grateful to her for her leadership and advice throughout the project. In addition, April Graham (Editorial Project Manager) very capably and efficiently supervised the development of the book, and Lisa Lamenzo (Project Manager) expertly coordinated its production.

The Editors of *Fundamental Neuroscience* hope that users of this book, and especially the students who will become the next generation of neuroscientists, find the subject matter of neuroscience as interesting and exciting as we do.

The Editors

This page intentionally left blank

About the Editors

Larry R. Squire is Distinguished Professor of Psychiatry, Neurosciences, and Psychology at the University of California School of Medicine, San Diego, and Research Career Scientist at the Veterans Affairs Medical Center, San Diego. He investigates the organization and neurological foundations of memory. He is a former President of the Society for Neuroscience and is a member of the National Academy of Sciences and the Institute of Medicine.

Darwin K. Berg is Distinguished Professor and Chair of Neurobiology in the Division of Biological Sciences at the University of California, San Diego. He has served as Chair of the Biology Department and Councilor of the Society for Neuroscience. He currently is Treasurer for the Society and a Board member of the Kavli Institute for Brain and Mind. His research is focused on the roles of nicotinic cholinergic signaling in the vertebrate nervous system.

Floyd E. Bloom is Professor Emeritus in the Molecular and Integrative Neuroscience Department (MIND) at The Scripps Research Institute. His recent awards include the Sarnat Award from the Institute of Medicine and the Salmon Medal of the New York Academy of Medicine. He is a former President of the Society for Neuroscience and is a member of the National Academy of Sciences and the Institute of Medicine.

Sascha du Lac is an Investigator of the Howard Hughes Medical Institute and an Associate Professor of Systems Neurobiology at the Salk Institute for Biological Studies. Her research focuses on the cellular machinery and circuit mechanisms that mediate cerebellar learning in the sense of balance.

Anirvan Ghosh is Stephen Kuffler Professor in the Division of Biological Sciences at the University of California, San Diego. His research interests include the development of synaptic connections in the central nervous system and the role of activity-dependent gene expression in cortical development. He is recipient of the Presidential Early Career Award for Scientists and Engineers, the Society for Neuroscience Young Investigator Award, and the UCSD Revelle College Outstanding Faculty Award.

Nicholas C. Spitzer is Distinguished Professor in the Division of Biological Sciences at the University of California, San Diego. His research is focused on the role of electrical activity in neuronal differentiation. He has been chairman of the Biology Department and the Neurobiology Section, a trustee of the Grass Foundation, and served as Councilor of the Society for Neuroscience. He is a member of the American Academy of Arts and Sciences, Co-Director of the Kavli Institute for Brain and Mind and Editor-in-Chief of BrainFacts.org.

This page intentionally left blank

Contributors

Igor Allaman Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Jocelyne Bachevalier Emory University, Atlanta, GA

Martha W. Bagnall Northwestern University, Chicago, IL

James F. Baker Northwestern University, Chicago, IL

Andrew H. Bell MRC Cognition and Brain Sciences Unit, Cambridge, UK

Floyd E. Bloom The Scripps Research Institute, La Jolla, CA

Scott T. Brady University of Illinois at Chicago, Chicago, IL

Marianne Bronner California Institute of Technology, Pasadena, CA

Peter Brophy University of Edinburgh, Edinburgh, Scotland

M. Christian Brown Harvard Medical School, Boston, MA

Elizabeth A. Buffalo Emory University School of Medicine, Atlanta, GA

Steven J. Burden NYU Medical Center, New York, NY

John H. Byrne University of Texas Medical School at Houston, Houston, TX

David N. Caplan Harvard Medical School, Boston, MA

J. Patrick Card University of Pittsburgh, Pittsburgh, PA

Hollis Cline The Scripps Research Institute, La Jolla, CA

Carol L. Colby University of Pittsburgh, Pittsburgh, PA

David R. Colman (Deceased.) Formerly at Montreal Neurological Institute, Montreal, Quebec, Canada

Jean de Vellis University of California at Los Angeles, Los Angeles, CA

Javier DeFelipe Cajal Institute, Madrid, Spain

Ariel Y. Deutch Vanderbilt University School of Medicine, Nashville, TN

Sascha du Lac The Salk Institute, La Jolla, CA

Gregory Elder The Mount Sinai School of Medicine and James J. Peters VA Medical Center, Bronx, NY

Patrick Emery University of Massachusetts Medical School, Worcester, MA

Barry J. Everitt University of Cambridge, Cambridge, United Kingdom

Jack L. Feldman University of California at Los Angeles, David Geffen School of Medicine, Los Angeles, CA

Miguel A. Gama Sosa Mount Sinai School of Medicine, New York, NY

Anirvan Ghosh University of California at San Diego, La Jolla, CA

Andrea C. Gore The University of Texas at Austin, Austin, TX

Jacqueline P. Gottlieb Columbia University College of Physicians and Surgeons, New York, NY

James L. Gould Princeton University, Princeton, NJ

Martyn Goulding The Salk Institute, La Jolla, CA

Sten Grillner Karolinska Institute, Stockholm, Sweden

William A. Harris Cambridge University, Cambridge, United Kingdom

Volker Hartenstein University of California at Los Angeles, Los Angeles, CA

Mary E. Hatten Rockefeller University, New York, NY

Michael Häusser University College London, London, United Kingdom

Stewart H. Hendry Johns Hopkins University, Baltimore, MD

J. Allan Hobson Harvard Medical School, Division of Sleep, Boston, MA

Patrick R. Hof Mount Sinai School of Medicine, New York, NY

Steven S. Hsiao Johns Hopkins University, Baltimore, MD

Yuh-Nung Jan Howard Hughes Medical Institute, University of California at San Francisco, San Francisco, CA

Jon H. Kaas Vanderbilt University, Nashville, TN

Sabine Kastner Princeton University, Princeton, NJ

Grahame Kidd The Cleveland Clinic Foundation, Cleveland, OH

Clemens Kiecker King's College London, London, United Kingdom

Eric I. Knudsen Stanford University School of Medicine, Stanford, CA

Christof Koch California Institute of Technology, Pasadena, CA and Allen Institute for Brain Science, Seattle, WA

Alex L. Kolodkin Johns Hopkins University School of Medicine, Baltimore, MD

George F. Koob The Scripps Research Institute, La Jolla, CA

Richard Krauzlis National Eye Institute, Bethesda, MD; The Salk Institute for Biological Studies, La Jolla, CA

Jeff Lichtman Harvard University, Cambridge, MA

Andrew Lumsden King's College London, London, United Kingdom

Pierre J. Magistretti Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland; University of Lausanne and CHUV, Lausanne, Switzerland

Joseph R. Manns Emory University, Atlanta, GA

Michael Mauk University of Texas at Austin, Austin, TX

David A. McCormick Yale University School of Medicine, New Haven, CT

Donald R. McCrimmon Northwestern University, Chicago, IL

George Z. Mentis Center for Motor Neuron Biology and Disease, Columbia University, New York, NY

Earl K. Miller Massachusetts Institute of Technology, Cambridge, MA

Carolanne E. Milligan Wake Forest School of Medicine, Winston-Salem, NC

Johnathan W. Mink University of Rochester Medical Center, Rochester, NY

Shaun F. Morrison Oregon Health and Science University, Portland, OR

Carl R. Olson Carnegie Mellon University, Pittsburgh, PA

Ronald W. Oppenheim Wake Forest School of Medicine, Winston-Salem, NC

Edward F. Pace-Schott Harvard Medical School, Boston, MA; Massachusetts General Hospital, Charlestown, MA

Luiz Pessoa University of Maryland, College Park, MD

Terry L. Powley Purdue University, West Lafayette, IN

Todd M. Preuss Emory University, Atlanta, GA

Peter R. Rapp National Institute on Aging, Baltimore, MD

R. Clay Reid Harvard Medical School, Boston, MA; Allen Institute for Brain Science, Seattle, WA

John H. Reynolds The Salk Institute, La Jolla, CA

Trevor W. Robbins University of Cambridge, Cambridge, United Kingdom

Josh Sanes Harvard University, Cambridge, MA

Joseph Santos-Sacchi Yale University School of Medicine, New Haven, CT

Peter Scheiffele Biozentrum of the University of Basel, Basel, Switzerland

Marc H. Schieber University of Rochester, Rochester, NY

Howard Schulman Allosteros Therapeutics, Sunnyvale, CA

Thomas L. Schwarz Children's Hospital, Boston, MA

Kristin Scott University of California at Berkeley, Berkeley, CA

Nelson Spruston HHMI Janelia Farm Research Campus, Ashburn, VA

Edward M. Stricker University of Pittsburgh, Pittsburgh, PA

Greg Stuart Australian National University, Canberra, Australia

Larry W. Swanson University of Southern California, Los Angeles, CA

Juan C. Tapia Columbia University, New York, NY

Marc Tessier-Lavigne The Rockefeller University, New York, NY

Roger B. H. Tootell Athinoula A. Martinos Center for Biomedical Imaging, Charlestown, MA; Harvard Medical School, Boston, MA

Bruce D. Trapp The Cleveland Clinic Foundation, Cleveland, OH

Leslie G. Ungerleider National Institute of Mental Health, Bethesda, MD

W. Martin Usrey Center for Neuroscience, University of California Davis, Davis, CA

Joseph G. Verbalis Georgetown University, Washington, DC

Christopher S. von Bartheld University of Nevada School of Medicine, Reno, NV

Jonathan D. Wallis University of California at Berkeley, Berkeley, CA

M. Neal Waxham University of Texas Medical School at Houston, Houston, TX

David R. Weaver University of Massachusetts Medical School, Worcester, MA

Stephen C. Woods University of Cincinnati Medical Center, Cincinnati, OH

SECTION I

FUNDAMENTAL NEUROSCIENCE

This page intentionally left blank

CHAPTER

1

Fundamentals of Neuroscience

A BRIEF HISTORY OF NEUROSCIENCE

The field of knowledge described in this book is neuroscience: the collected multidisciplinary sciences that analyze the nervous system to understand the biological basis for behavior. Modern studies of the nervous system have been ongoing since the middle of the nineteenth century. Neuroanatomists studied the brain's shape, its cellular structure, and its circuitry; neurochemists studied the brain's chemical composition, its lipids and proteins; neurophysiologists studied the brain's bioelectric properties; and psychologists and neuropsychologists investigated the organization and neural substrates of behavior and cognition.

In contrast to these separate investigative pursuits, the term *neuroscience* was introduced in the mid-1960s to signal the beginning of an era in which each of these disciplines would work together cooperatively, sharing a common language, common concepts, and a common goal: to understand the structures and functions of both the normal and abnormal brain. Neuroscience today spans a wide range of research endeavors from the molecular biology of nerve cells (i.e., the genes encoding the proteins needed for nervous system function) to the biological basis of normal and disordered behavior, emotion, and cognition (i.e., the mental properties by which individuals interact with one another and their environments). For a more complete, but concise, history of the neurosciences, see Kandel and Squire (2000).

Neuroscience is currently one of the most rapidly growing areas of science. Indeed, the brain is sometimes referred to as "the last frontier of biology." In 1971, 1100 scientists convened at the first annual meeting of the Society for Neuroscience. In 2010, 31,500 scientists participated at the society's 40th annual meeting, at which more than 15,000 research presentations were made.

THE TERMINOLOGY OF NERVOUS SYSTEMS IS HIERARCHICAL, DISTRIBUTED, DESCRIPTIVE, AND HISTORICALLY BASED

Beginning students of neuroscience may understandably find themselves confused. Nervous systems of many organisms have their cell assemblies and macroscopically visible components named by multiple overlapping and often synonymous terms. With a necessarily gracious view to the past, this confusing terminology could be viewed as the intellectual cost of focused discourse with predecessors in the enterprise. The nervous systems of invertebrate organisms often are designated for their spatially directed collections of neurons responsible for local control of operations, such as the thoracic or abdominal ganglia, which receive sensations and direct motoric responses for specific body segments, all under the general control of a cephalic ganglion whose role includes sensing the external environment.

In vertebrates, the components of the nervous system were named for both their appearance and their location. As noted by Swanson & Bota (2010), and expanded upon in Chapter 2 of this volume, the names of the major parts of the brain were based on creative interpretations of early dissectors of the brain, attributing names to brain segments based on their appearance in the freshly dissected state: hippocampus (shaped like the sea horse) or amygdala (shaped like the almond), cerebrum (the main brain), and cerebellum (a small brain).

© 2013 Elsevier Inc. All rights reserved.

NEURONS AND GLIA ARE CELLULAR BUILDING BLOCKS OF THE NERVOUS SYSTEM

This book lays out our current understanding in each of the important domains that together define the full scope of modern neuroscience. The structure and function of the brain and spinal cord are most appropriately understood from the perspective of their highly specialized cells: the neurons, the interconnected, highly differentiated, bioelectrically driven, cellular units of the nervous system, and their more numerous support cells, the glia. Given the importance of these cellular building blocks in all that follows, a brief overview of their properties may be helpful.

Neurons Are Heterogeneously Shaped, Highly Active Secretory Cells

Neurons are classified in many different ways: according to function (sensory, motor, or interneuron), location (cortical, spinal, etc.), the identity of the transmitter they synthesize and release (glutamatergic, cholinergic, etc.), and their shape (pyramidal, granule, mitral, etc.). Microscopic analysis focuses on their general shape and, in particular, the number of extensions from the cell body. Most neurons have one axon, often branched, to transmit signals to interconnected target neurons. Other processes, termed dendrites, extend from the nerve cell body (also termed the perikaryon—the cytoplasm surrounding the nucleus of the neuron) to receive synaptic contacts from other neurons; dendrites may branch in extremely complex patterns and may possess multiple short protrusions called dendritic spines. Neurons exhibit the cytological characteristics of highly active secretory cells with large nuclei; large amounts of smooth and rough endoplasmic reticulum; and frequent clusters of specialized smooth endoplasmic reticulum (Golgi apparatus), in which secretory products of the cell are packaged into membrane-bound organelles for transport out of the cell body proper to the axon or dendrites. Neurons and their cellular extensions are rich in microtubules—elongated tubules approximately 24 nm in diameter. Microtubules support the elongated axons and dendrites and assist in the reciprocal transport of essential macromolecules and organelles between the cell body and the distant axon or dendrites.

Neurons Communicate Chemically through Specialized Contact Zones

The sites of interneuronal communication in the central nervous system (CNS) are termed *synapses* in the CNS and *junctions* in somatic, motor, and autonomic nervous systems. Paramembranous deposits of specific proteins essential for transmitter release, response, and catabolism characterize synapses and junctions morphologically. These specialized sites are presumed to be the active zone for transmitter release and response. Paramembranous proteins constitute a specialized junctional adherence zone, termed the ***synaptolemma***. Like peripheral junctions, central synapses also are denoted by accumulations of tiny (500 to 1500 Å) organelles, termed synaptic vesicles. The proteins of these vesicles have been shown to have specific roles in transmitter storage; vesicle docking onto presynaptic membranes, voltage- and Ca^{2+}-dependent secretion, and the recycling and restorage of previously released transmitter molecules.

Synaptic Relationships Fall into Several Structural Categories

Synaptic arrangements in the CNS fall into a wide variety of morphological and functional forms that are specific for the neurons involved. The most common arrangement, typical of hierarchical pathways, is either the ***axodendritic*** or the ***axosomatic*** synapse, in which the axons of the cell of origin make their functional contact with the dendrites or cell body of the target neuron, respectively. A second category of synaptic arrangement is more rare: forms of functional contact between adjacent cell bodies (***somasomatic***) and overlapping dendrites (***dendrodendritic***). Within the spinal cord and some other fields of ***neuropil*** (relatively acellular areas of synaptic connections), serial ***axoaxonic*** synapses are relatively frequent. Here, the axon of an interneuron ends on the terminal of a long-distance neuron as that terminal contacts a dendrite, or on the segment of the axon that is immediately distal to the soma, termed the initial segment, where action potentials arise. Many presynaptic axons contain local collections of typical synaptic vesicles with no opposed specialized synaptolemma. These are termed ***boutons en passant***. The release of a transmitter may not always occur at such sites.

Synaptic Relationships also Belong to Diverse Functional Categories

As with their structural representations, the qualities of synaptic transmission can also be functionally categorized in terms of the nature of the neurotransmitter that provides the signaling; the nature of the receptor molecule on the postsynaptic neuron, gland, or muscle; and the mechanisms by which the postsynaptic cell transduces the neurotransmitter signal into transmembrane changes. So-called "fast" or "classical" neurotransmission is the functional variety seen at the vast majority of synaptic and junctional sites, with a rapid

onset and a rapid ending, generally employing excitatory amino acids (glutamate or aspartate) or inhibitory amino acids (γ-aminobutyrate, GABA, or glycine) as the transmitter. The effects of those signals are largely attributable to changes in postsynaptic membrane permeability to specific cations or anions and the resulting depolarization or hyperpolarization, respectively. Other neurotransmitters, such as the monoamines (dopamine, norepinephrine, serotonin) and many neuropeptides, produce changes in excitability that are much more enduring. Here the receptors activate metabolic processes within the postsynaptic cells—frequently to add or remove phosphate groups from key intracellular proteins; multiple complex forms of enduring postsynaptic metabolic actions are under investigation. The brain's richness of signaling possibilities comes from the interplay on common postsynaptic neurons of these multiple chemical signals.

The Operative Processes of Nervous Systems are also Hierarchical

As research progressed, it became clear that neuronal functions could best be fitted into nervous system function by considering their operations at four fundamental hierarchical levels: molecular, cellular, systems, and behavioral. These levels rest on the fundamental principle that neurons communicate chemically, by the activity-dependent secretion of neurotransmitters, at specialized points of contact named synapses.

At the ***molecular*** level of operations, the emphasis is on the interaction of molecules—typically proteins that regulate transcription of genes, their translation into proteins, and their posttranslational processing. Proteins that mediate the intracellular processes of transmitter synthesis, storage, and release, or the intracellular consequences of intercellular synaptic signaling, are essential neuronal molecular functions. Such transductive molecular mechanisms include the neurotransmitters' receptors, as well as the auxiliary molecules that allow these receptors to influence the short-term biology of responsive neurons (through regulation of ion channels) and their longer-term regulation (through alterations in gene expression). Completion of the human, chimpanzee, rat, and mouse genomes can be viewed as an extensive inventory of these molecular elements, more than half of which are thought to be either highly enriched in the brain or even exclusively expressed there. Genes unique to human brains have been reported and appear to be most frequently expressed during the brain's development. Recent progress has made clear that the ***introns*** (the sections of the genome that contain DNA code that does not lead to production of specific proteins; these noncoding sequences are the largest part of the genome) are important functional control elements that regulate which genes are expressed and which transcribed messenger RNAs are translated into proteins.

At the ***cellular*** level of neuroscience, the emphasis is on interactions between neurons through their synaptic transactions and between neurons and glia. Much current cellular-level research focuses on the biochemical systems within specific cells that mediate such phenomena as pacemakers for the generation of circadian rhythms or that can account for activity-dependent adaptation. Research at the cellular level strives to determine which specific neurons and which of their most proximate synaptic connections may mediate a behavior or the behavioral effects of a given experimental perturbation.

At the ***systems*** level, emphasis is on the spatially distributed sensors and effectors that integrate the body's response to environmental challenges. There are sensory systems, which include specialized senses for hearing, seeing, feeling, tasting, and balancing the body. Similarly, there are motor systems for trunk, limb, and fine finger motions and internal regulatory systems for visceral regulation (e.g., control of body temperature, cardiovascular function, appetite, and salt and water balance). These systems operate through relatively sequential linkages, and interruption of any link can destroy the function of the system.

Systems-level research also includes research into cellular systems that innervate the widely distributed neuronal elements of the sensory, motor, or visceral systems, such as the pontine neurons with highly branched axons that innervate diencephalic, cortical, and spinal neurons. Among the best studied of these divergent systems are the monoaminergic neurons, which have been linked to the regulation of many behavioral outputs of the brain, ranging from feeding, drinking, thermoregulation, and sexual behavior. Monoaminergic neurons also have been linked to such higher functions as pleasure, reinforcement, attention, motivation, memory, and learning. Dysfunctions of these systems have been hypothesized as the basis for some psychiatric and neurological diseases, supported by evidence that medications aimed at presumed monoamine regulation provide useful therapy.

At the ***behavioral*** level of neuroscience research, emphasis is on the interactions between individuals and their collective environment. Research at the behavioral level centers on the integrative phenomena that link populations of neurons (often operationally or empirically defined) into extended specialized circuits, ensembles, or more pervasively distributed "systems" that integrate the physiological expression of a learned, reflexive, or spontaneously generated behavioral response. Behavioral research also includes the operations of higher mental activity, such as memory, learning, speech, abstract reasoning, and consciousness. Conceptually, "animal models" of human psychiatric diseases are based

on the assumption that scientists can appropriately infer from observations of behavior and physiology (heart rate, respiration, locomotion, etc.) that the states experienced by animals are equivalent to the emotional states experienced by humans expressing these same sorts of physiological changes.

As the neuroscientific bases for some elemental behaviors have become better understood, new aspects of neuroscience applied to problems of daily life have begun to emerge. Methods for the noninvasive detection of activity in certain small regions of the cerebral cortex can now resolve areas that are active simultaneously and presumed to act together in networks of cortical assemblies. These concurrently active networks have been linked to discrete forms of mental activity. Such advances have given rise to the concept that it is possible to understand where in the brain the decision-making process occurs or to identify the kinds of information necessary to decide whether or not to act. The detailed quantitative data that now exist on the details of neuronal structure, function, and behavior have driven the development of computational neurosciences. This new branch of neuroscience research seeks to predict the performance of neurons, neuronal properties, and neural networks based on their discernible quantitative properties.

Some Principles of Brain Organization and Function

The central nervous system is most commonly divided into major structural units, consisting of the major physical subdivisions of the brain. Thus, mammalian neuroscientists divide the central nervous system into the brain and spinal cord and further divide the brain into regions readily seen by the simplest of dissections. Based on research that has demonstrated that these large spatial elements derive from independent structures in the developing brain, these subdivisions are well accepted. Mammalian brain thusly is divided into hindbrain, midbrain, and forebrain, each of which has multiple highly specialized regions within it. In deference to the major differences in body structure, invertebrate nervous systems most often are organized by body segment (cephalic, thoracic, abdominal) and by anterior–posterior placement.

Neurons within the vertebrate CNS operate either within layered structures (such as the olfactory bulb, cerebral cortex, hippocampal formation, and cerebellum) or in clustered groupings (the defined collections of central neurons, which aggregate into "nuclei" in the central nervous system and into "ganglia" in the peripheral nervous system, and in invertebrate nervous systems). The specific connections between neurons within or across the macro-divisions of the brain are essential to the brain's functions. It is through their patterns of neuronal circuitry that individual neurons form functional ensembles to regulate the flow of information within and between the regions of the brain.

CELLULAR ORGANIZATION OF THE BRAIN

Present understanding of the cellular organization of the CNS can be viewed simplistically according to three main patterns of neuronal connectivity.

Three Basic Patterns of Neuronal Circuitry Exist

Long hierarchical neurons typically constitute the primary sensory and motor pathways. Here the transmission of information is highly sequential, and interconnected neurons are related to one another in a hierarchical fashion. Primary receptors (in the retina, inner ear, olfactory epithelium, tongue, or skin) transmit first to primary relay cells, then to secondary relay cells, and finally to the primary sensory fields of the cerebral cortex. For motor output systems, the reverse sequence exists with impulses descending hierarchically from the motor cortex to the spinal motoneuron. It is at the level of the motor and sensory systems that beginning scholars of the nervous system will begin to appreciate the complexities of neuronal circuitry by which widely separated neurons communicate selectively. This hierarchical scheme of organization provides for a precise flow of information, but such organization suffers the disadvantage that destruction of any link incapacitates the entire system.

Local circuit neurons establish their connections mainly within their immediate vicinity. Such local circuit neurons frequently are small and may have relatively few processes. Interneurons expand or constrain the flow of information within their small spatial domain and may do so without generating action potentials, given their short axons.

Single-source divergent neurons operate within the hypothalamus, pons, and medulla. From their clustered nuclear locations, these neurons extend multiple branched and divergent connections to many target cells, almost all of which lie outside the brain region in which the neurons of origin are located. Neurons with divergent circuitry could be considered more as interregional interneurons rather than as sequential elements within any known hierarchical system. For example, different neurons of the noradrenergic nucleus, the locus coeruleus (named for its blue pigmented color in primate brains), project from the pons to either the cerebellum, spinal cord, hypothalamus, or

several cortical zones to modulate synaptic operations within those regions.

Glia Are Supportive Cells to Neurons

Neurons are not the only cells in the CNS. According to most estimates, neurons are outnumbered, perhaps by an order of magnitude, by the various nonneuronal supportive cellular elements. Nonneuronal cells include macroglia, microglia, and cells of the brain's blood vessels, cells of the choroid plexus that secrete the cerebrospinal fluid, and meninges, sheets of connective tissue that cover the surface of the brain and comprise the cerebrospinal fluid-containing envelope that protects the brain within the skull.

Macroglia are the most abundant supportive cells; some are categorized as ***astrocytes*** (nonneuronal cells interposed between the vasculature and the neurons, often surrounding individual compartments of synaptic complexes). Astrocytes play a variety of metabolic support roles, including furnishing energy intermediates and providing for the supplementary removal of excessive extracellular neurotransmitter secretions (see Chapter 13). A second prominent category of macroglia is the myelin-producing cells, the oligodendroglia. Myelin, made up of multiple layers of their compacted membranes, insulates segments of long axons bioelectrically and accelerates action potential conduction velocity. Microglia are relatively uncharacterized supportive cells believed to be of mesodermal origin and related to the macrophage/monocyte lineage. Some microglia reside quiescently within the brain. During periods of intracerebral inflammation (e.g., infection, certain degenerative diseases, or traumatic injury), circulating macrophages and other white blood cells are recruited into the brain by endothelial signals to remove necrotic tissue or to defend against the microbial infection.

The Blood–Brain Barrier Protects Against Inappropriate Signals

The blood–brain barrier is an important permeability barrier to selected molecules between the bloodstream and the CNS. Evidence of a barrier is provided by the greatly diminished rate of access of most lipophobic chemicals between plasma and brain; specific energy-dependent transporter systems permit selected access. Diffusional barriers retard the movement of substances from brain to blood as well as from blood to brain. The brain clears metabolites of transmitters into the cerebrospinal fluid by excretion through the acid transport system of the choroid plexus. The blood–brain barrier is much more permeable in the hypothalamus and in several small, specialized organs (termed circumventricular organs; see Chapters 34 and 39) lining the third and fourth ventricles of the brain: the median eminence, area postrema, pineal gland, subfornical organ, and subcommissural organ. The peripheral nervous system (e.g., sensory and autonomic nerves and ganglia) has no such diffusional barrier.

The Central Nervous System Can Initiate Limited Responses to Damage

Because neurons of the CNS are terminally differentiated cells, they cannot undergo proliferative responses to damage, as can cells of skin, muscle, bone, and blood vessels. Nevertheless, small populations of previously unrecognized stem cells have been shown capable of undergoing regulated proliferation and differentiation into neurons and astrocytes. These stem cells, more prominent in rodent than primate brains, provide a natural means for selected neuronal replacement in some regions of the nervous system. As a result, neurons have evolved other adaptive mechanisms to provide for the adaptive responses to environmental demands and can be brought into play for maintenance of function following injury. These adaptive mechanisms range from activity-dependent regulation of gene expression, to modification of synaptic structure, function, and can include actual localized axonal sprouting, changes in dendritic spine numbers and shapes, and new synapse creation. These adaptive mechanisms endow the brain with considerable capacity for structural and functional modification well into adulthood. This plasticity is not only considered to be activity dependent but also to be reversible with disuse. Plasticity is pronounced within the sensory systems (see Chapter 23) and is quite prominent in the motor systems as well. The molecular mechanisms employed in memory and learning may rely upon very similar processes as those involved in structural and functional plasticity.

ORGANIZATION OF THIS TEXT

With these overview principles in place, which are detailed more extensively in Section II, we can resume our preview of this book. Another major domain of our field is nervous system development (Section III). How does a simple epithelium differentiate into specialized collections of cells and ultimately into distinct brain structures? How do neurons grow processes that find appropriate targets some distance away? How do nascent neuronal activity and embryonic experience shape activity?

Sensory systems and motor systems (Sections IV and V) encompass how the nervous system receives information

from the external world and how movements and actions are produced (e.g., eye movements and limb movements). These questions range from the molecular level (how are odorants, photons, and sounds transduced into informative patterns of neural activity?) to the systems and behavioral level (which brain structures control eye movements and what are the computations required by each structure?).

An evolutionarily old function of the nervous system is to regulate respiration, heart rate, sleep and waking cycles, food and water intake, and hormones to maintain internal homeostasis and to permit daily and longer reproductive cycles. In this area of regulatory systems (Section VI), we explore how organisms remain in balance with their environment, ensuring that they obtain the energy resources needed to survive and reproduce. At the level of cells and molecules, the study of regulatory systems concerns the receptors and signaling pathways by which particular hormones or neurotransmitters prepare the organism to sleep, to cope with acute stress, or to seek food or reproduce. At the level of brain systems, we ask such questions as what occurs in brain circuitry to produce thirst or to create a self-destructive problem such as drug abuse?

In recent years, the disciplines of psychology and biology have increasingly found common ground, and this convergence of psychology and biology defines the modern topics of behavioral and cognitive neuroscience (Section VII). These topics concern the so-called higher mental functions: perception, attention, language, memory, thinking, and the ability to navigate in space. Work on these problems traditionally has drawn on the techniques of neuroanatomy, neurophysiology, neuropharmacology, and behavioral analysis. More recently, behavioral and cognitive neuroscience has benefited from several new approaches: the use of computers to perform detailed formal analyses of how brain systems operate and how cognition is organized; noninvasive neuroimaging techniques, such as positron emission tomography and functional magnetic resonance imaging, to obtain dynamic images of the living human brain in action; and molecular biological methods, such as single gene knockouts in mice, which can relate genes to brain systems and to behavior. An especially attractive area of current investigation is the use of fluorescent proteins, such as enhanced green fluorescent protein, to demarcate individual neuronal trajectories, visualizing the "connectome," the collection of all neuronal circuits. An offshoot of this approach is to create transgenic animals in which neurons expressing specific neurotransmitter receptors can be visually identified in living brain slices for precise investigation of synaptic properties. A third example, termed optogenetics, causes light-sensitive proteins to be expressed in discrete neuronal circuits to be activated or inhibited in intact behaving experimental animals, including nonhuman primates.

THIS BOOK IS INTENDED FOR A BROAD RANGE OF SCHOLARS OF THE NEUROSCIENCES

This textbook is for anyone interested in neuroscience. In preparing it we have focused primarily on graduate students just entering the field, understanding that some of you will have majored in biology, some in psychology, some in mathematics or engineering, and even some, like me, in German literature. It is hoped that through the text, the explanatory boxes, and, in some cases, the supplementary readings, you will find the book to be both understandable and enlightening. In many cases, advanced undergraduate students will find this book useful as well.

Medical students may find that they need additional clinical correlations that are not provided here. However, it is hoped that most medical scholars at least will be able to use our textbook in conjunction with more clinically oriented material. Finally, to those who have completed their formal education, it is hoped that this text can provide you with some useful information and challenging perspectives, whether you are active neuroscientists wishing to learn about areas of the field other than your own or individuals who wish to enter neuroscience from a different area of inquiry. We invite all of you to join us in the adventure of studying the nervous system.

CLINICAL ISSUES IN THE NEUROSCIENCES

Many fields of clinical medicine are directly concerned with the brain. The branches of medicine tied most closely to neuroscience are neurology (the study of the diseases of the brain), neurosurgery (the study of the surgical treatment of neurological disease), and psychiatry (the study of behavioral, emotional, and mental diseases). Other fields of medicine also make important contributions, including radiology (the use of radiation for such purposes as imaging the brain—initially with x-rays and, more recently, with positron emitters and magnetic waves) and pathology (the study of pathological tissue). To make connections to the many facets of medicine that are relevant to neuroscience, this book includes discussion of a limited number of clinical conditions in the context of basic knowledge in neuroscience.

The Spirit of Exploration Continues

Less than a decade into the twenty-first century, the Hubble space telescope continues to transmit information about the uncharted regions of the universe and

clues to the origin of the cosmos. This same spirit of adventure also is being directed to the most complex structure that exists in the universe: the human brain. The complexity of the human brain is enormous, describable only in astronomical terms. For example, the number of neurons in the human brain (about 10^{12} or 1000 billion) is approximately equal to the number of stars in our Milky Way galaxy. Whereas the possibility of understanding such a complex device is certainly daunting, it is nevertheless true that an enormous amount has already been learned. The promise and excitement of research on the nervous system have captured the attention of thousands of students and working scientists. What is at stake is not only the possibility of discovering how the brain works. It is estimated that diseases of the brain, including both neurological and psychiatric illnesses, affect as many as 50 million individuals annually in the United States alone, at an estimated societal cost of hundreds of billions of dollars in clinical care and lost productivity. The prevention, treatment, and cure of these diseases will ultimately be found through neuroscience research. Moreover, many of the issues currently challenging societies globally—instability within the family, illiteracy, poverty, and violence, as well as improved individualized programs of education—could be illuminated by a better understanding of the brain.

THE GENOMIC INVENTORY IS A GIANT STEP FORWARD

Possibly the single largest event in the history of biomedical research was presented in published form in February 2001: the initial inventory of the human genome. By using advanced versions of the powerful methods of molecular biology, several large scientific teams have been able to take apart all of an individual's human DNA in very refined ways, amplify the amounts of the pieces, determine the order of the nucleic acid bases in each of the fragments, and then put those fragments back together again across the 23 pairs of human chromosomes.

Having determined the sequences of the nucleic acids, it was possible to train computers to read the sequence information and spot the specific signals that identify the beginning and ending of sequences likely to encode proteins. Furthermore, the computer systems could then sort those proteins by similarity of sequences (motifs) within their amino acid building blocks. After sorting, the computers could next assign the genes and gene products to families of similar proteins whose functions had already been established. In this way, scientists were rapidly able to predict approximately how many proteins could be encoded by the genome (all of the genes an individual has). Whole genome data are now available for humans, for some nonhuman primates, for rats, and for mice.

Scientifically, this state of information has been termed a "draft" because it is based on a very dense, but not quite complete, sample of the whole genome. What has been determined still contains a very large number of interruptions and gaps. Some of the smaller genes, whose beginning and ending are most certain, could be thought of as parts in a reassembled Greek urn, held in place by bits of blank clay until further excavation is done. However, having even this draft has provided some important realities.

Similar routines allowed these genomic scholars to determine how many of those mammalian genes were like genes we have already recognized in the smaller genomes of other organisms mapped out previously (yeast, worm (*Caenorhabditis elegans*), and fruit fly (*Drosophila melanogaster*)) and how many other gene forms may not have been encountered previously. Based on current estimates, it would appear that despite the very large number of nucleotides in the human and other mammalian genomes, about 30 times the length of the worms and more like 15 times the fruit fly, mammals may have only twice as many genes—perhaps some 30,000 to 40,000 altogether. Compared to other completed genomes, the human genome has greatly increased its representation of genes related to nervous system function, tissue-specific developmental mechanisms, and immune function and blood coagulation. Importantly for diseases of the nervous system that are characterized by the premature death of neurons, there appears to have been a major expansion in the numbers of genes related to initiating the process of intentional cell death, or apoptosis. Although still controversial, genes regulating primate brain size have been reported, but links to intellectual capacity remain unproven.

Two major future vistas can be imagined. To create organisms as complex as humans from relatively so few genes probably means that the richness of the required proteins is based on their modifications, either during transcription of the gene or after translation of the intermediate messenger RNA into the protein. These essential aspects of certain proteins account for a small number of brain diseases that can be linked to mutations in a single gene, such as Huntington's disease (see Chapter 31). Second, though compiling this draft inventory represents a stunning technical achievement, there remains the enormously daunting task of determining, for example, where in the brain's circuits specific genes normally are expressed and how that expression pattern may be altered by the demands of illness or an unfriendly environment. That task, at present, is one for which there are as yet no tools equivalently as powerful as those used to acquire the flood of sequence

data with which we are now faced. This stage has been referred to as the end of "naive reductionism."

In the fall of 2005, a six-nation consortium of molecular biologists announced the next phase of genomic research. The new focus will be toward refining the initial inventories to compare whole genomes of healthy and affected individuals for a variety of complex genetic illnesses (the HapMap project). Complex genetic diseases, such as diabetes mellitus, hypertension, asthma, depression, schizophrenia, and alcoholism, arise through the interactions of multiple short gene mutations that can increase or decrease one's vulnerability to a specific disease depending on individual life experiences. Ultimately, as the speeds of genome sequencing improve still further and the cost is reduced, it may be possible to predict what diseases will be more likely to affect a given person and to predict lifestyle changes that person could undertake to improve his or her opportunities to remain healthy. Given the recently recognized importance of the intronic sequences of the genome in determining which genes are expressed, so-called "Next-Gen" or "deep sequencing" methods of very high speed and resolution are forcing investigators to re-examine the genetic basis of heritable disorders. Increasing attention is being devoted to epigenetic regulation of gene expression by means of chemical modification of gene sequences and their expression.

In order to benefit from the enormously rich potential mother lode of genetic information, next we must determine where these genes are expressed, what functions they can control, and what sorts of controls other gene products can exert over them. In the nervous system, where cell–cell interaction is the main operating system in relating molecular events to functional behavioral events, discovering the still murky properties of activity-dependent gene expression will require enormous investment.

NEUROSCIENCE TODAY: A COMMUNAL ENDEAVOR

As scientists, we draw from the work of those who came before us, using other scientists' work as a foundation for our own. We build on and extend previous observations and, it is hoped, contribute something to those who will come after us. The information presented in this book is the culmination of hundreds of years of research. To help acquaint you with some of this work, we have described many of the key experiments of neuroscience throughout the book. We also have listed some of the classic papers of neuroscience and related fields at the end of each chapter, and we invite you to read some of them for yourselves.

The pursuit of science has not always been a communal endeavor. Initially, research was conducted in relative isolation. The scientific "community" that existed at the time consisted of intellectuals who shared the same general interests, terminology, and paradigms. For the most part, scientists were reluctant to collaborate or share their ideas broadly, because an adequate system for establishing priority for discoveries did not exist. However, with the emergence of scientific journals in 1665, scientists began disseminating their results and ideas more broadly because the publication record could be used as proof of priority. Science then began to progress much more rapidly, as each layer of new information provided a higher foundation on which new studies could be built.

Gradually, an interactive community of scientists evolved, providing many of the benefits that contemporary scientists enjoy: working as part of a community allows for greater specialization and efficiency of effort. This not only allows scientists to study a topic in greater depth but also enables teams of researchers to attack problems from multidisciplinary perspectives. The rapid feedback and support provided by the community help scientists refine their ideas and maintain their motivation. It is this interdependence across space and time that gives science much of its power.

With interdependence, however, comes vulnerability. In science, as in most communities, codes of acceptable conduct have evolved in an attempt to protect the rights of individuals while maximizing the benefits they receive. Some of these guidelines are concerned with the manner in which research is conducted, and other guidelines refer to the conduct of scientists and their interactions within the scientific community. Let us begin by examining how new knowledge is created.

THE CREATION OF KNOWLEDGE

Over the years, a generally accepted procedure for conducting research has evolved. This process involves examining the existing literature, identifying an important question, and formulating a research plan. Often, new experimental pathways are launched when one scientist reads with skepticism the observations and interpretations of another and decides to test their validity. Sometimes, especially at the beginning of a new series of experiments, the research plan is purely "descriptive"—for example, determining the structure of a protein or the distribution of a neurotransmitter in the brain. Descriptive initial research is essential to the subsequent inductive phase of experimentation, the movement from observations to theory, seasoned with wisdom and curiosity. Descriptive experiments are valuable both because of the questions that they attempt

to answer and because of the questions that their results allow us to ask. Information obtained from descriptive experiments provides a base of knowledge on which a scientist may draw to develop hypotheses about cause and effect in the phenomenon under investigation. For example, once we identify the distribution of a particular transmitter within the brain or the course of a pathway of connections through descriptive work, we may then be able to develop a theory about what function that transmitter or pathway serves.

Once a hypothesis has been developed, the researcher then has the task of designing and performing experiments that are likely to disprove that hypothesis if it is incorrect. This is referred to as the deductive phase of experimentation, the movement from theory to observation. Through this paradigm the neuroscientist seeks to narrow down the vast range of alternative explanations for a given phenomenon. Only after attempting to disprove the hypothesis as thoroughly as possible can scientists be adequately assured that their hypothesis is a plausible explanation for the phenomenon under investigation.

A key point in this argument is that data may only lend support to a hypothesis rather than provide absolute proof of its validity. In part, this is because the constraints of time, money, and technology allow a scientist to test a particular hypothesis only under a limited set of conditions. Variability and random chance may also contribute to the experimental results. Consequently, at the end of an experiment, scientists generally report only that there is a statistical probability that the effect measured was due to intervention rather than to chance or variability.

Given that one can never prove a hypothesis, how do "facts" arise? At the conclusion of their experiments, the researchers' first task is to report their findings to the scientific community. The dissemination of research findings often begins with an informal presentation at a laboratory or departmental meeting, eventually followed by presentation at a scientific meeting that permits the rapid exchange of information more broadly. One or more research articles published in peer-reviewed journals ultimately follow the verbal communications. Such publications are not simply a means to allow the authors to advance as professionals (although they are important in that respect as well). Publication is an essential component of the advancement of science. As we have already stated, science depends on sharing information, replicating and thereby validating experiments, and then moving forward to solve the next problem. Indeed, a scientific experiment, no matter how spectacular the results, is not completed until the results are published. More likely, publication of "spectacular" results will provoke a skeptical scientist into doing an even more telling experiment, and knowledge will evolve.

RESPONSIBLE CONDUCT

Although individuals or small groups may perform experiments, new knowledge is ultimately the product of the larger community. Inherent in such a system is the need to be able to trust the work of other scientists—to trust their integrity in conducting and reporting research. Thus, it is not surprising that much emphasis is placed on the responsible conduct of research.

Research ethics encompasses a broad spectrum of behaviors. Where one draws the line between sloppy science and unethical conduct is a source of much debate within the scientific community. Some acts are considered to be so egregious that despite personal differences in defining what constitutes ethical behavior, the community generally recognizes certain research practices as behaviors that are unethical. These unambiguously improper activities consist of fabrication, falsification, and plagiarism. Fabrication refers to making up data, falsification is defined as altering data, and plagiarism consists of using another person's ideas, words, or data without attribution. Each of these acts significantly harms the scientific community.

Fabrication and falsification in a research paper taint the published literature by undermining its integrity. Not only is the information contained in such papers misleading in itself, but other scientists may unwittingly use that information as the foundation for new research. If, when reported, these subsequent studies cite the previous, fraudulent publication, the literature is further corrupted. Thus, through a domino-like effect, one paper may have a broad negative impact on the scientific literature. Moreover, when fraud is discovered, a retraction of the paper provides only a limited solution, as there is no guarantee that individuals who read the original article will see the retraction. Given the impact that just one fraudulent paper may have, it is not surprising that the integrity of published literature is a primary ethical concern for scientists.

Plagiarism is also a major ethical infraction. Scientific publications provide a mechanism for establishing priority for a discovery. As such, they form the currency by which scientists earn academic positions, gain research grants to support their research, attract students, and receive promotions. Plagiarism denies the original author of credit for his or her work. This hurts everyone: the creative scientist is robbed of credit, the scientific community is hurt by the disincentive to share ideas and research results, and the individual who has plagiarized—like the person who has fabricated or falsified data—may well find his or her career ruined.

In addition to the serious improprieties just described, which are in fact extremely rare, a variety of much more frequently committed "misdemeanors" in

the conduct of research can also affect the scientific community. Like fabrication, falsification, and plagiarism, some of these actions are considered to be unethical because they violate a fundamental value, such as honesty. For example, most active scientists believe that honorary authorship—listing as an author someone who did not make an intellectual contribution to the work—is unethical because it misrepresents the origin of the research. In contrast, other unethical behaviors violate standards that the scientific community has adopted. For example, although it is generally understood that material submitted to a peer-reviewed journal as part of a research manuscript has never been published previously and is not under consideration by another journal, instances of retraction for dual publications can be found on occasion.

Scientific Misconduct Has Been Formally Defined by U.S. Governmental Agencies

The serious misdeeds of fabrication, falsification, and plagiarism generally are recognized throughout the scientific community. These were broadly recognized by federal regulations in 1999 as a uniform standard of scientific misconduct by all agencies funding research. What constitutes a misdemeanor is less clear, however, because variations in the definitions of accepted practices are common. There are several sources of this variation. Because responsible conduct is based in part on conventions adopted by a field, it follows that there are differences among disciplines with regard to what is considered to be appropriate behavior. For example, students in neuroscience usually coauthor papers with their advisor, who typically works closely with them on their research. In contrast, students in the humanities often publish papers on their own even if their advisor has made a substantial intellectual contribution to the work reported. Within a discipline, the definition of acceptable practices may also vary from country to country. Because of animal use regulations, neuroscientists in the United Kingdom do relatively little experimental work with animals on the important topic of stress, whereas in the United States this topic is seen as an appropriate area of study so long as guidelines are followed to ensure that discomfort to the animals is minimized.

The definition of responsible conduct may change over time. For example, some protocols that were once performed on human and animal subjects may no longer be considered ethical. Indeed, ethics evolve alongside knowledge. We may not currently be able to know all of the risks involved in a procedure, but as new risks are identified (or previously identified risks refuted), we must be willing to reconsider the facts and adjust our policies as necessary. In sum, what is considered to be ethical behavior may not always be obvious, and therefore we must actively examine what is expected of us as scientists.

Having determined what is acceptable practice, we then must be vigilant. Each day neuroscientists are faced with a number of decisions having ethical implications, most of them at the level of misdemeanors: Should a data point be excluded because the apparatus might have malfunctioned? Have all the appropriate references been cited and are all the authors appropriate? Might the graphic representation of data mislead the viewer? Are research funds being used efficiently? Although individually these decisions may not significantly affect the practice of science, cumulatively they can exert a great effect.

In addition to being concerned about the integrity of the published literature, we must be concerned with our public image. Despite concerns over the level of federal funding for research, neuroscientists are among the privileged few who have much of their work funded by taxpayer dollars. Highly publicized scandals damage the public image of our profession and hurt all of us who are dependent on continued public support for our work. They also reduce the public credibility of science and thereby lessen the impact that we can expect our findings to have. Thus, for our own good and that of our colleagues, the scientific community, and the public at large, we must strive to act with integrity.

SUMMARY

You are about to embark on a tour of fundamental neuroscience. Enjoy the descriptions of the current state of knowledge, read the summaries of some of the classic experiments on which that information is based, and consult the references that the authors have drawn on to prepare their chapters. Think also about the ethical dimensions of the science you are studying; your success as a professional and the future of our field depend on it.

References

Boyden, E. S., Katoh, A., Pyle, J. L., Chatila, T. A., Tsien, R. W., & Raymond, J. L. (2006). Selective engagement of plasticity mechanisms for motor memory storage. *Neuron, 51,* 823–834.

Kandel, E. R., & Squire, L. R. (2000). Neuroscience: Breaking down scientific barriers to the study of brain and mind. *Science, 290,* 1113–1120.

Suggested Readings

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience, 28,* 403–450.

Cherniak, C. (1990). The bounded brain: Toward quantitative neuroanatomy. *Journal of Cognitive Neuroscience, 2,* 58–68.

Committee on the Conduct of Science. (2009). *On being a scientist* (3rd ed.). Washington, DC: National Academy Press, National Academy of Sciences.

Cowan, W. M., & Kandel, E. R. (2001). Prospects for neurology and psychiatry. *JAMA, 285,* 594–600.

Day, R. A. (1994). *How to write and publish a scientific paper* (4th ed.). Phoenix, AZ: Oryx Press.

Dent, E. W., Merriam, E. B., & Hu, X. (2010). The dynamic cytoskeleton: Backbone of dendritic spine plasticity. *Current Opinion in Neurobiology, 20,* 494–502.

Gräff, J., Kim, D., Dobbin, M. M., & Tsai, L. H. (2011). Epigenetic regulation of gene expression in physiological and pathological brain processes. *Physiological Reviews, 91,* 603–649.

Greengard, P. (2001). The neurobiology of slow synaptic transmission. *Science, 294,* 1024–1030.

Kuhn, T. S. (1996). *The structure of scientific revolutions* (3rd ed.). Chicago: Univ. of Chicago Press.

Raichle, M., & Mintun, M. A. (2006). Brain work and brain imaging. *Annual Review of Neuroscience, 29,* 449–476.

Shepherd, G. M. (2003). *The synaptic organization of the brain*. New York: Oxford Uni. Press.

Swanson, L. W., & Bota, M. (2010). Foundational model of structural connectivity in the nervous system with a schema for wiring diagrams, connectome, and basic plan architecture. *Proceedings of the National Academy of Sciences of the United States of America, 107,* 20610–20617.

Floyd E. Bloom

This page intentionally left blank

CHAPTER

2

Basic Plan of the Nervous System

INTRODUCTION

The brain often is compared with a computer these days. True, the brain is a computer, but it is a very special kind of computer—a biological computer that has evolved by natural selection over hundreds of millions of years and countless generations. Furthermore, it has no obvious design features in common with human-engineered computers. Instead, the brain is a unique organ that thinks and feels, generates behavioral interactions with the environment, keeps bodily physiology relatively stable, and enables reproduction of the species—its most important role from evolution's grand perspective. And for strictly personal reasons the brain is the most precious thing we have simply because it is the organ of consciousness, as reflected in René Descartes's famous seventeenth-century aphorism, "I think, therefore I am."

Aristotle first emphasized that structure and function are inextricably intertwined, two sides of the same coin, with structure providing obvious physical constraints on function. Just think about the difference between a hammer and a saw. Unfortunately, as knowledge becomes more and more specialized, there is a tendency to analyze the structure, function, and chemistry of the nervous system from different, sometimes even isolated, perspectives. The main theme of this chapter is the basic structure-function organization of the nervous system: what are the parts and how are they interconnected into functional systems? In other words, what are the organizing principles—the basic design features—of its circuitry?

Evolution and development are two approaches often used to understand biological complexity—because they start with the simplest condition—and the human brain is far and away the most complex object we know of, with its roughly 100 billion neurons and 100 trillion axonal connections between them. One remarkable conclusion emerging from these two perspectives is that nerve cells in all animals—from jellyfish to humans—are basically the same in terms of cell biology; what changes most during ontogeny and phylogeny is the *arrangement* of nerve cells into functional circuits: the architecture of the nervous system. The "bricks" or "Legos" are similar, but the "buildings" constructed with them can vary tremendously in size and functionality. An ultimate goal of neuroscience may be to understand the human brain, but remarkable progress can nevertheless be made through analyzing "lower" animals and early embryos. The other equally remarkable conclusion is that all vertebrates, from fish to humans, share a common basic plan of the nervous system, with the same major parts and functional systems.

EVOLUTION HIGHLIGHTS: GENERAL ORGANIZING PRINCIPLES

Protists and the simplest multicellular animals (sponges) display ingestive, defensive, reproductive, and other behaviors without any nervous system whatsoever, raising the question, what is the adaptive value of adding a nervous system to an organism? We will now examine key structure-function correlates of nervous system organization in animals with relatively simple body plans and behaviors.

The Nerve Net is the Simplest Type of Nervous System

In his provocative 1919 book *The Elementary Nervous System*, George Parker outlined a reasonable scenario for how nervous systems evolved. An updated version begins with the first multicellular animals—similar to modern-day sponges—that emerged over half a billion years ago. They are seemingly amorphous animals that spend their adult

© 2013 Elsevier Inc. All rights reserved.

lives immobile, submerged in water. Their relatively simple behavior is mediated largely by a set of primitive smooth muscle cell (myocyte) sphincters regulating water flow through body wall pores. These specialized cells are called *independent effectors* because their contraction is evoked by stimuli like stretch or environmental chemicals acting directly on the plasma membrane of individual cells.

The first animal phylum with a nervous system was the Cnidaria, which includes jellyfish, corals, sea anemones, and the elegantly simple hydra. In contrast to sponges, hydra locomote and show active feeding behavior (Fig. 2.1). These behaviors are coordinated and mediated by a nervous system, a network of specialized units or cells called nerve cells or *neurons* (Box 2.1).

Sensory Neurons

Hydra's body wall is simple, with an outer *ectoderm* layer contacting the external environment, an inner *endoderm* layer facing the body cavity's internal environment and promoting digestion and waste elimination, and a vague middle or *meso* layer in between. Neurons probably differentiated initially from the ectoderm, and perhaps the first to evolve were *sensory neurons*. One cytoplasmic extension of these bipolar cells facing the external environment became specialized to detect stimuli much weaker than those activating independent effectors, whereas the other pole became specialized to transmit information about these stimuli to a *group* of independent effectors (Fig. 2.3). Experimental evidence

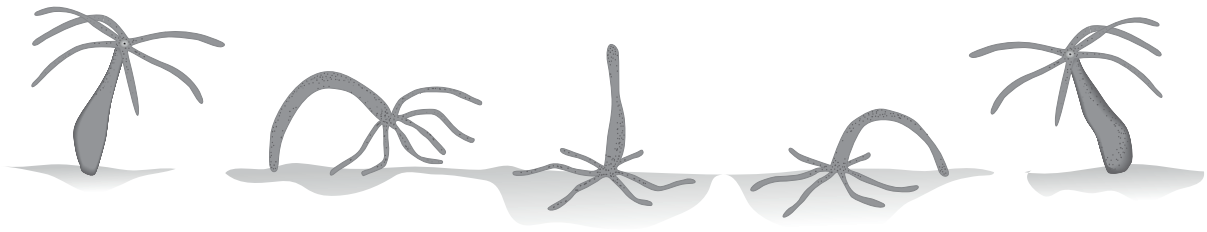


FIGURE 2.1 Locomotor behavior in hydra resembles a series of somersaults, as shown in the sequence beginning on the left. The tiny black dot in the region between the tentacles in the figure at the far right is the animal's mouth. Ingestive (feeding) behavior involves guiding food particles into the mouth with coordinated tentacle movements.

BOX 2.1

THE NEURON DOCTRINE

The cell theory, which states that all organisms are composed of individual cells, was developed around the middle of the nineteenth century by Mattias Schleiden and Theodor Schwann. However, this unitary vision of the cellular nature of life was not immediately applied to the nervous system, as most biologists at the time believed in the cytoplasmic continuity of nervous system cells. Later in the century, the most prominent advocate of this *reticularist* view was Camillo Golgi, who proposed that axons entering the spinal cord actually fuse with other axons (Fig. 2.2A). The reticularist view was challenged most thoroughly by Santiago Ramón y Cajal, a founder of contemporary neuroscience and without doubt the greatest observer of neuronal architecture. In beautifully written and carefully reasoned deductive arguments, Cajal presented us with what is now known as the *neuron doctrine*. This great concept in essence states that the cell theory applies to the nervous system: each neuron is an individual entity, the basic unit of neural circuitry (Fig. 2.2B). The acrimonious debate between reticularists and proponents of the neuron doctrine raged for decades. Over the years, the validity of the neuron doctrine has been supported by a wealth of accumulated data. Nevertheless, the reticularist view is not entirely incorrect, because some neurons do act syncytially via specialized intercellular gap junctions, a feature that is more prominent during embryogenesis.

In 1897, Charles Sherrington postulated that neurons establish functional contact with one another and with other cell types via a theoretical structure he called the synapse (Greek *synaptein*, to fasten together). It was not until 50 years later that the structural existence of synapses was demonstrated by electron microscopy (see Fig. 3.3). The synaptic complex is built around an *adhesive* junction, and in this and other respects the complex is quite similar to the desmosome and the adherens junctions of epithelia. In fact, similarities in ultrastructure between the adherens junction and the synaptic complex of central nervous tissue were noted even in early electron microscopic studies (see Peters, Palay, & Webster, 1991).

indicates that sensory neurons provide four major selective advantages in evolution:

- Increased stimulus sensitivity
- Faster effector cell responses
- Stronger behavioral responses because multiple effector cells are influenced
- Sensory neurons responding to different stimulus modalities can be distributed strategically in different body regions

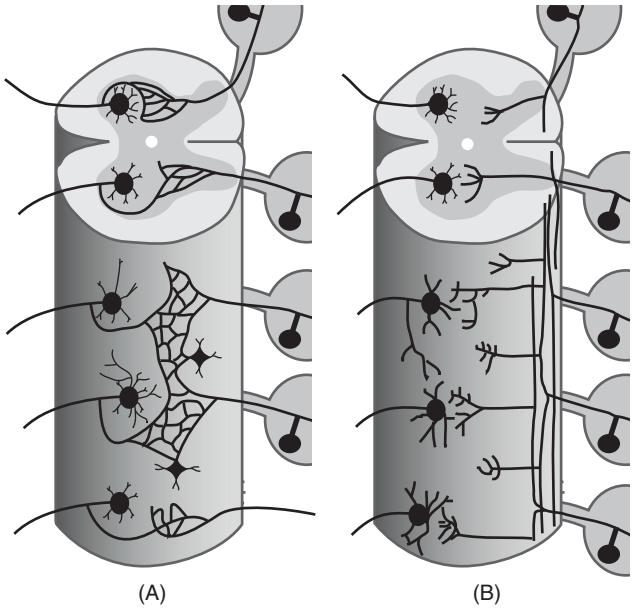


FIGURE 2.2 Two competing views: The nervous system as a reticulum or the neuron doctrine. (A) Proponents of the reticular theory believed that neurons are physically continuous with one another, forming an uninterrupted network. (B) In contrast, the neuron doctrine regards each neuron as an individual entity communicating with target cells by way of contiguity rather than continuity, across an appropriate intercellular gap. *Adapted from Cajal (1909–1911).*

The bipolar shape of sensory neurons is fundamentally important. The prototypical theory about neural circuit organization was presented by Santiago Ramón y Cajal in his classic "bible" of structural neuroscience, *The Histology of the Nervous System in Man and Vertebrates* (1909–1911). According to the cornerstone *functional polarity theory*, information normally flows in one direction through most neurons, and thus through most neural circuits—from *dendrites* and *cell body*, the input or receptive parts of the neuron, to a single *axon*, the output or effector part. In other words, most neurons have two classes of *cytoplasmic extensions*: one or more dendrites detecting inputs and a single axon conducting an output that can influence multiple cells through branching or *collateralization*. At least in early developmental stages, all sensory neurons have this fundamental bipolar shape, and over the course of evolution they have become specialized to detect a remarkable variety of stimuli from light, temperature, and a wide range of chemicals and ions, to vibration and other mechanical deformations.

Motor Neurons

A second stage of differentiation or complexity in hydra's nervous system was the addition of neurons between sensory and effector. They are defined as motor neurons (*motoneurons*) because they directly innervate effector cells (usually muscle or gland cells), which in turn receive their inputs from sensory neurons (Figs. 2.2B, 2.3C). Conceptually, this provides a *two-layered nervous system*, with the first or top layer having sensory neurons, and the second or bottom layer having motor neurons. In this prototypical network sensory neurons project (send axon collaterals) to multiple motoneurons, and then each motoneuron innervates a set of effector cells (with a motoneuron and its effector cell set defined as a *motor unit*). During

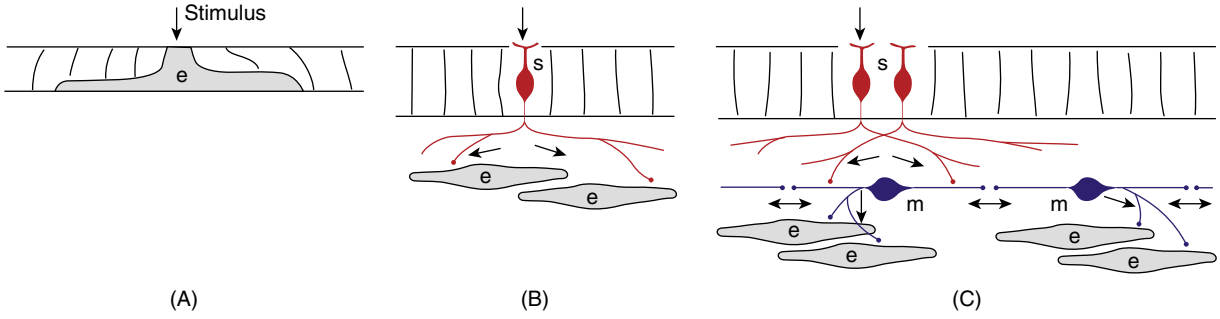


FIGURE 2.3 Activation of effector cells (e) in simple animals. (A) Sponges lack a nervous system; stimuli act directly on effector cells, which are thus called independent effectors. (B) In cniderians, bipolar sensory neurons (s) differentiate in the ectoderm. The sensory neuron outer extension detects stimuli and is thus a dendrite. The inner extension of some sensory neurons transmits information directly to effector cells and is thus an axon. Because this type of sensory neuron innervates effector cells directly, it is actually a sensorimotor neuron. (C) Most cnidarian sensory neurons send their axon to motoneurons (m), which in turn send an axon to effector cells. Cnidarian motoneurons may also have lateral extensions interacting with other motoneurons, and these extensions typically conduct information in either direction via reciprocal chemical synapses (and are thus amacrine extensions). Arrows show the direction of information flow.

an animal's normal behavior, information flow is unidirectional or polarized from one cell type, sensory neuron, to another cell type, motoneuron, to a third cell type, effector. This is the basic definition of a simple reflex, as defined by Charles Sherrington in his cornerstone of systems neuroscience, *The Integrative Action of the Nervous System* (1906).

In this hypothetical scenario (Fig. 2.3) an environmental stimulus detected by a sensory neuron's dendrite is transmitted by its axon to the dendrites of a motoneuron population. Then the axon of each motoneuron innervates an effector cell population. This is the functional polarity rule applied to a simple two-layer, sensory-motor network mediating reflex behavior.

Another general feature of the hydra two-layered nervous system has been observed: sensory neurons do not innervate one another, whereas motoneurons do interact directly. Here, motoneurons have two projection classes: one to effector cells and another to other motoneurons. Structurally and functionally, many of these hydra motoneurons also have two types of output extensions. One is a typical axon innervating an effector cell population. However, the other is an extension that contacts homologous extensions from other motoneurons. Interestingly, many of these "tangential" extensions between motoneurons transmit information in either direction; either motoneuron can transmit information to the other via these extensions. This is an exception to the functional polarity rule and is mediated by reciprocal rather than the more common unidirectional synapses. Cajal (1909–1911) described several examples of neurons that lack a clear axon (in retina, olfactory bulb, and intestine) and called them amacrine cells. As an extension of this, it is useful to divide neuronal extensions into three types: dendritic (input), axonal (output), or amacrine (bidirectional).

Adding a second layer to the nervous system has obvious adaptive advantages related to increased capacity for response complexity and integration. Consider a stimulus to one specific part of the animal or even one sensory neuron. Its influence may radiate to distant parts of the animal because one sensory neuron innervates multiple motoneurons, those motoneurons innervate additional motoneurons, and each motoneuron innervates multiple effector cells—an example of what Cajal called "avalanche conduction." There may be great *divergence* between stimulus and effector cells producing a response, with the actual divergence pattern shaped by the structure–function architecture of the nervous system: how the neurons and their interconnections are arranged in the body. It is easy to imagine how this arrangement in hydra might coordinate the tentacles to bring a food morsel detected by just one of them to the mouth or how it might coordinate locomotion (Fig. 2.1).

A second basic consequence of this structural arrangement is information *convergence* in the nervous system. Just consider a particular motoneuron: it can receive inputs from more than one sensory neuron and from other motoneurons as well.

Nerve Nets

At first glance, hydra's nervous system is distributed fairly uniformly around the radially symmetrical body wall and tentacles (Fig. 2.4). Its essentially double-layered arrangement of distributed sensory and motor neurons is called a *nerve net*. However, in certain regions of the body with specialized function, like around the mouth and base of the tentacles, neurons tend to aggregate—a tendency toward *centralization* that will now be examined more carefully.

Bilateral Symmetry, Centralization, and Cephalization Emerge in Flatworms

In contrast to cnidarians, flatworms are bilaterally symmetrical predators with rostral (head) and caudal (tail) ends, and dorsal and ventral surfaces. These changes in body plan and behavior are accompanied by equally important changes in nervous system organization. Many flatworm neurons are clustered into distinct ganglia interconnected by longitudinal and transverse axon bundles called *nerve cords* (Fig. 2.5). This condensation of neural elements, or centralization, allows faster

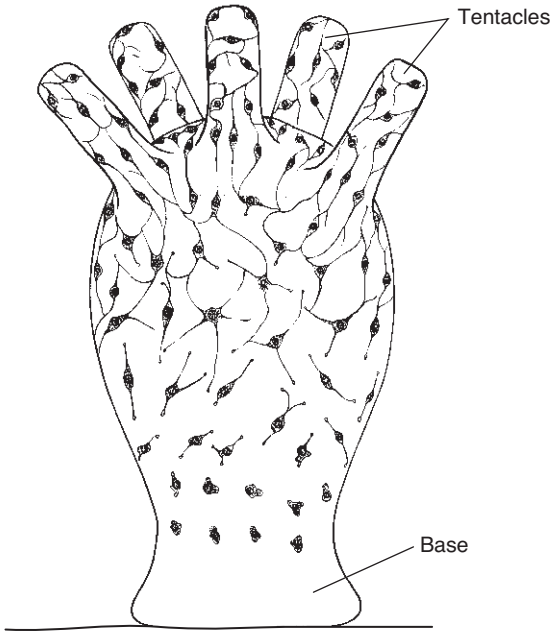


FIGURE 2.4 The nerve net of hydra, a simple cnidarian, is spread diffusely throughout the body wall of the animal. This drawing shows maturation of the nerve net in a hydra bud, starting near the base and finishing near the tentacles. *Refer to McConnell (1932) and Koizumi (2002).*

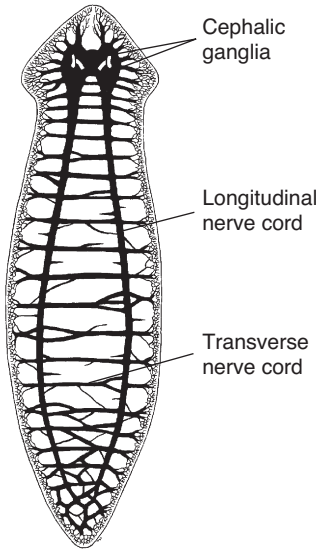


FIGURE 2.5 The nervous system of the planarian, a flatworm, includes longitudinal and transverse nerve cords associated with centralization, and two fused cephalic ganglia (the invertebrate brain) in the rostral end associated with cephalization. Centralization and cephalization probably are related to the flatworm's bilateral symmetry and ability to swim forward rapidly. *Refer to Lentz (1968). Reproduced with permission from Yale University Press.*

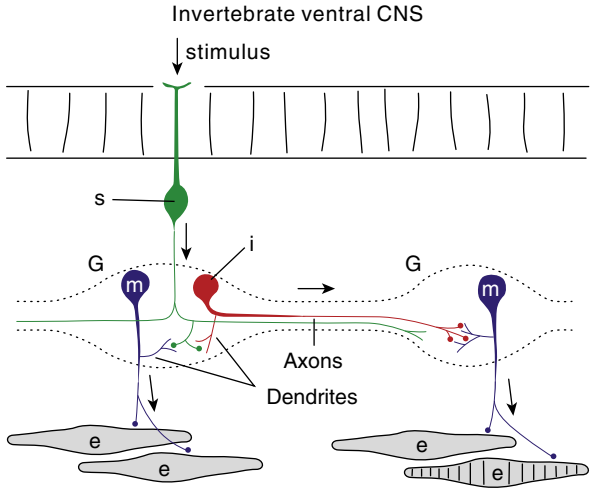


FIGURE 2.6 Invertebrate ganglia (G) usually display two neuron classes: motor neurons (m) and interneurons (i), both typically unipolar, with dendrites arising from a single axon. Here neuronal cell bodies are arranged peripherally and synapses occur in a central region called the neuropil. Sensory neurons (s) usually innervate motoneurons and interneurons but not effectors (e). Arrows show the usual direction of information flow.

and thus more efficient communication between neurons because cellular material is conserved and conduction times are reduced. The largest, most complex ganglia (*cephalic ganglia* or invertebrate brains) are localized rostrally, where they receive information from specialized sensory receptors in the front of the animal as it swims. Bilateral symmetry, centralization, and cephalization are three cardinal organizational trends in nervous system evolution.

Interneurons

Flatworms are the simplest animals, with an abundant, clearly distinct third neuron division, *interneurons*, which are interpolated between sensory and motor neurons (Fig. 2.6). As already noted, Cajal recognized some atypical interneurons that apparently lack distinguishable dendrites and axon (amacrine neurons, or more precisely, amacrine interneurons). However, most interneurons have recognizable dendrites and axon and so presumably transmit information down the axon in only one direction: toward its terminals. They are *typical neurons* conforming to the functional polarity rule.

One consequence of adding a third "layer" of neurons to the nervous system is simply to increase convergence and divergence of information processing, and thus the capacity for response complexity. There are, however, three other critical functions interneurons subserve. They can act as *excitatory* or *inhibitory* "*switches*" in neuronal networks, assemblies of them can act as *pattern detectors* and *generators* between sensory and motor neurons, and they can be *pacemakers* if they generate intrinsic rhythmical activity patterns.

By this definition the vast majority of vertebrate brain neurons are interneurons. So it is useful at the outset to recognize two broad interneuron categories: intraregional or local and interregional or projection. *Intraregional interneurons*, or local circuit neurons, have an axon that remains confined to a distinguishable gray matter region or ganglion, whereas *interregional interneurons* send a longer axon to a different gray matter region or ganglion, although it may also generate local axon collaterals.

The omnidirectional information flow typical of cnidarian nerve nets is unusual in the rest of the animal kingdom, where most neurons are functionally polarized with information flowing through neural circuits sequentially from dendrites and cell body to axon and axon terminals. However, most invertebrate motoneurons and interneurons are unipolar: a single extension, the axon, emanates from the cell body. Dendrites branch *from* the axons in the center of a ganglion—entering the *neuropil*—where most synapses are formed (Fig. 2.6). In vertebrates most neurons are multipolar, with several dendrites, plus an axon extending from the cell body or a dendrite.

Features of simple nervous systems are preserved throughout evolution. For example, the part of the nervous system in the wall of the human gastrointestinal tract (the enteric plexuses) has many features of a highly refined nerve net, and a "layer" of amacrine

BOX 2.2

CAJAL: ICONOCLAST TO ICON

Santiago Ramón y Cajal (1852–1934) is considered by many people to be the founder of modern neuroscience—a peer of Darwin and Pasteur in nineteenth-century biology. He was born in the tiny Spanish village of Petilla de Aragon on May 1, 1852, and as related in his delightful autobiography, he was somewhat mischievous as a child and determined to become an artist, much to the consternation of his father, a respected local physician. However, he eventually entered the University of Zaragoza and received a medical degree in 1873. As a professor of anatomy at Zaragoza, his interests were mostly in bacteriology (the nineteenth-century equivalent of molecular biology today in terms of an exciting biomedical frontier) until 1887, when he visited Madrid at age 35 and first saw through the microscope histological sections of brain tissue treated with the Golgi method, which had been introduced in 1873. Although very few workers had used this technique, Cajal saw immediately that it offered great hope in solving the most vexing problem of nineteenth-century neuroscience: How do adult nerve cells interact with one another? This realization galvanized and directed the rest of his scientific life, which was extremely productive in terms of originality, scope, and accuracy.

Shortly after Jacob Schleiden, Theodor Schwann, and Rudolf Virchow proposed the cell theory in the late 1830s, Joseph von Gerlach Sr. and Otto Deiters suggested that nerve tissue was special in the sense that nerve cells are not independent units but instead form a continuous cytoplasmic syncytium or reticular net (Fig. 2.2A). This concept was later refined by Camillo Golgi, who, based on the use of his silver chromate method, concluded that axons of nerve cells form a continuous reticular net, whereas in contrast dendrites do not anastomose but instead serve a nutritive role, much like the roots of a tree. Using the same technique, Cajal almost immediately arrived at the opposite conclusion, based first on his examination of the cerebellum, and later of virtually all other parts of the nervous system. In short, he proposed that neurons interact by way of contact or contiguity rather than by continuity and are thus structurally independent units, which was finally proven when the electron microscope was used in the 1950s to characterize synapses. This concept became known as the *neuron doctrine*.

Cajal's second major conceptual achievement was the theory of **functional polarity**, which stated that the dendrites and cell bodies of neurons receive information, whereas the single axon with its collaterals transmits information to the other cells. This rule allows prediction of information flow direction through neural circuits based on the morphology or shape of individual neurons forming them, and it was the cornerstone of Charles Sherrington's (1906) revolutionary physiological analysis of mammalian reflex organization. Recent evidence that many dendrites transmit an action potential or graded potential in the retrograde direction would not violate the tenets of the functional polarity theory unless the potential led to altered membrane potentials in the associated presynaptic axon—and if this were the case the "dendrite" would be classed instead as an amacrine process (see text).

Around the close of the nineteenth century, Cajal made a remarkable series of discoveries at the cellular level. In addition to the two concepts outlined earlier, they include (1) the mode of axon termination in the adult CNS (1888), (2) the dendritic spine (1888), (3) the first diagrams of reflex pathways based on the neuron doctrine and functional polarity (1890), (4) the axonal growth cone (1890), (5) the chemotactic theory of synapse specificity (1892), and (6) the hypothesis that learning could be based on the selective strengthening of synapses (1895).

In one of the great ironies in the history of neuroscience, Cajal and Golgi shared the Nobel Prize for Medicine in 1906, though they had used the same technique to elaborate fundamentally different views on nervous system organization! The meeting in Stockholm may not have diminished the great personal friction between them. In 1931 Cajal wrote, "What a cruel irony of fate to pair like Siamese twins united by the shoulders, scientific adversaries of such contrasting characters."

interneurons is found in the human retina and olfactory bulb.

A Segmented Ventral Nerve Cord Typifies Annelids and Arthropods

Annelid worms and arthropods have even more complex body plans and behaviors than flatworms, partly because of *segmentation*. Body segments (*metameres*) are repeated serially along the body's rostrocaudal axis and presumably share a common underlying genetic developmental program, although terminal differentiation (adult structure) may vary in each. This strategy allows for more complex body plans (including the nervous system) to evolve without a linear or exponential increase in genetic material. Annelids and all the more complex invertebrates share another characteristic feature, a *ventral nerve cord* with a pair of ganglia (or a single fused

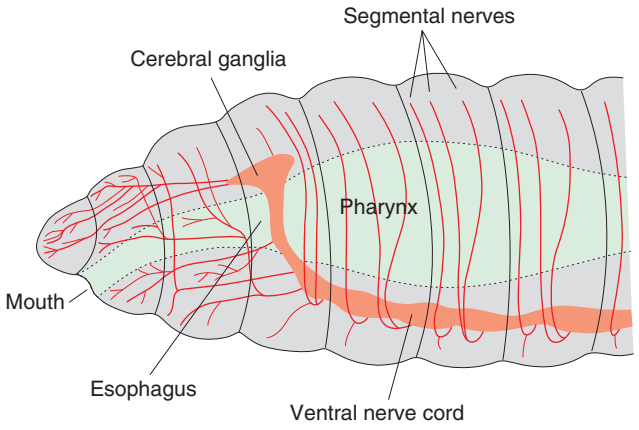


FIGURE 2.7 Nervous system organization in the rostral end of an annelid worm. A ventral nerve cord with more or less distinct ganglia connects with a fused pair of cerebral ganglia (brain) dorsal to the pharynx (part of the digestive tract). Note nerves extending from ventral nerve cord and cerebral ganglia. *Refer to Brusca and Brusca (1990).*

ganglion) in each segment, and longitudinal axon bundles between ganglia in adjacent segments (Fig. 2.7). Transverse *nerves* also extend from each ganglion to sensory structures and muscles in the same segment.

The Basic Plan of the Vertebrate Nervous System is Found in Lancelets

Vertebrates are a subphylum of the Chordates and are the most complex of all animals in terms of structure and behavior. They share a basic body plan where common organ systems are arranged in a relatively strict anatomical relationship with one another (Box 2.3 and Figs. 2.8, 2.9). Like other Chordates, vertebrates display two key features during some part of their life: a cartilaginous rod, the *notochord*, extending dorsally along the body, and above it a hollow *dorsal nerve cord*. In most vertebrates the notochord's body stiffening and protective functions are supplanted by the vertebral column and bony skull, with the notochord reduced to a series of cartilaginous cushions (discs) between or within the vertebrae. The vertebrate central nerve cord is tremendously expanded, thickened, and folded to form the brain and spinal cord (the central nervous system, CNS).

The vertebrate nervous system's basic parts are revealed in the lancelet (amphioxus), a simple, nonvertebrate chordate (subphylum Cephalochordata). The lancelet is a slender, fish-like filter-feeder living half buried in the sand of shallow, tropical marine waters (Fig. 2.9). The body is stiffened by a notochord, and a hollow dorsal nerve cord runs the length of the body, generating segmental nerves innervating muscles and organs. Locomotor behavior (swimming) is produced by alternately contracting right and left segmental muscles (myotomes). Without a notochord, these contractions would shorten the animal rather than generate forward propulsive force.

Although typical vertebrate brain regions are not obvious rostrally in the lancelet nerve cord, genes specifying

BOX 2.3

ANATOMICAL RELATIONSHIPS IN THE VERTEBRATE BODY

To describe the physical relationships between structures in the nervous system and the rest of the vertebrate body, it is best to use terms that accurately and unambiguously describe the position of a given structure in three dimensions without reference to the external world. The major axis of the body is the **rostrocaudal axis**, which extends along the length of the animal from the *rostrum* (beak) to the *cauda* (tail) (Fig. 2.8), as well as the length of the embryonic neural plate and neural tube (see Figs. 2.10 and 2.12). A second axis, the orthogonal **dorsoventral axis**, is vertical and runs from the *dorsum* (back) to the *ventrum* (belly). Finally, the third perpendicular axis, the **mediolateral axis**, is horizontal and runs from the midline (medial) to the lateral margin of the animal (lateral). Unfortunately, the rostrocaudal axis undergoes complex bending during embryogenesis, and the bending pattern is unique to each species. It would be ideal if the three cardinal axes were used in a topologically accurate way—say, with reference to the body as it might appear with a "straightened out" rostrocaudal axis. In practice, however, this is rarely the case, which leads to a certain degree of ambiguity, as is obvious when looking at the fish, frog, cat, and human bodies shown in Figure 2.8.

The problem is especially difficult in human anatomy where use of an idiosyncratic terminology has a long, ingrained tradition. The basic principles are much easier to illustrate than to describe in writing (see Fig. 2.8), but one major source of confusion in the human brain is related to the fact that the rostrocaudal axis makes a 90-degree bend in the midbrain region (unlike in rodents and carnivores, for example, where the axis is relatively straight). The other source of confusion is simply the different names that are used. For example, in human anatomy the spinal cord has anterior and posterior horns, and posterior root ganglia, whereas in other mammals they usually are referred to as ventral and dorsal horns, and dorsal root ganglia. The merits of a uniformly applied nomenclature based on comparative structural principles seem obvious.

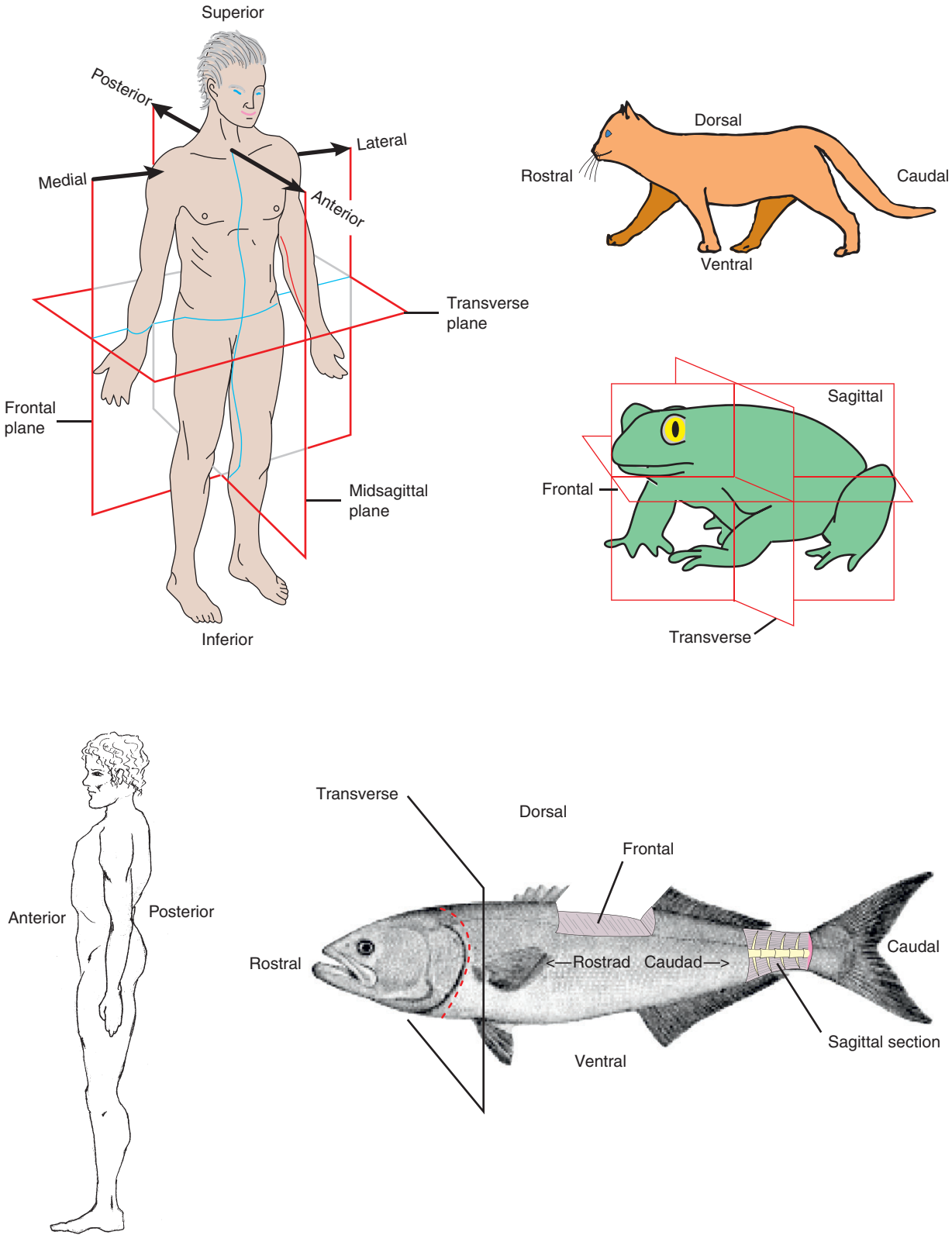


FIGURE 2.8 Orientation of the vertebrate body. Orientation planes for fish, quadrupeds, and bipeds are depicted. Associated with the three cardinal planes (rostrocaudal, dorsoventral, and mediolateral) are three orthogonal planes: frontal, sagittal, and transverse, which are the same in all early vertebrate embryos. For more explanation, see Williams (1995).

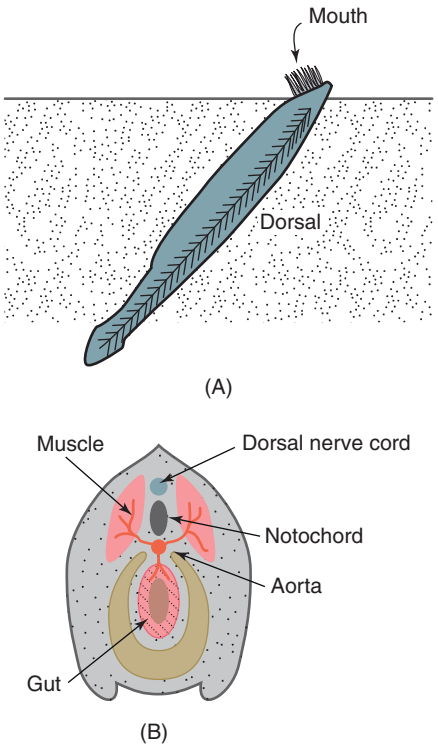


FIGURE 2.9 The lancelet (amphioxus) is a forerunner of the vertebrates. (A) Lateral view of the animal in its native habitat under the ocean floor, with its mouth protruding above the sand. (B) A cross-section of the lancelet body showing relationships between gut, aorta, notochord, and dorsal nerve cord. *Adapted from Cartmill, Hylander, and Shafland (1987).*

early vertebrate head embryogenesis also are expressed rostrally in the lancelet body. Thus, some components of the molecular program specifying modern vertebrate head development apparently were present early in chordate evolution (Holland & Takahashi, 2005).

Summary

The cniderian nerve net displays most of the basic cellular features of nervous system organization, including convergence and divergence of sensory and motor information. In more complex bilaterally symmetrical invertebrates, neurons and axons tend to aggregate in ganglia, nerve cords, and nerves (centralization), and there is a greater concentration of neurons and sensory organs in the body's rostral end (cephalization). Segmented invertebrates have a ventral nerve cord that includes a bilateral pair of ganglia (or single fused ganglion) in each segment. The primitive chordate, lancelet, displays the basic nervous system organization characteristic of vertebrates, including mammals and humans.

DEVELOPMENT REVEALS BASIC VERTEBRATE PARTS

One nineteenth-century biology triumph was the demonstration that early stages of embryogenesis are fundamentally the same in all vertebrates. The CNS and heart are the first organs to differentiate in the embryo, and the basic CNS divisions differentiating early in development are also common to all vertebrates. The names and arrangement of these divisions are the starting point for regional or topographic neuroanatomical nomenclature (Swanson, 2000a).

Nervous System Regionalization Begins in the Neural Plate

During embryogenesis the CNS develops as a hollow cylinder (*neural tube*) from a topologically flat sheet of cells (*neural plate*), by a process of *neurulation* (Chapter 13). Here we simply consider macroscopic structural changes during the transformation.

The neural plate is a spoon-shaped differentiation of the trilaminar embryonic disc's one-cell-thick ectodermal layer (Fig. 2.10). Its wide end lies rostrally and becomes the brain, whereas the narrow end lies caudally and becomes the spinal cord—the two major CNS divisions. A midline *neural groove* divides the neural plate into right and left halves, so the plate displays three cardinal morphogenetic features: *polarity*, *bilateral symmetry*, and *regionalization*. Furthermore, the neural plate differentiates from rostral to caudal, so the brain plate regionalizes first. Signs of this include appearance of the *optic vesicles*, evaginating near the rostral end of the neural plate (in the presumptive hypothalamus); a midline *infundibulum* evaginating between optic vesicles and rostral end of the notochord, and indicating the presumptive pituitary stalk (again in presumptive hypothalamus); and the *otic rhombomere* (presumptive or future rhombomere 4), a swelling near the center of what will become the adult rhombicbrain (Figs. 2.11, left; 2.14).

At the junction between neural plate and remaining ectoderm (later forming the skin's epidermal layer or skin) lies a narrow strip of transitional ectoderm, the *neural crest*, a distinctive vertebrate feature (Fig. 2.10). It generates a variety of adult structures, including most neurons of the peripheral nervous system (PNS).

In summary, the CNS and PNS divisions are represented by the neural plate and neural crest, respectively, during the neural plate stage of vertebrate development. The two major CNS divisions, the brain and spinal cord, are also indicated in the neural plate, which at this developmental stage is topologically simple: a bilaterally symmetrical, flat sheet that is one cell thick.

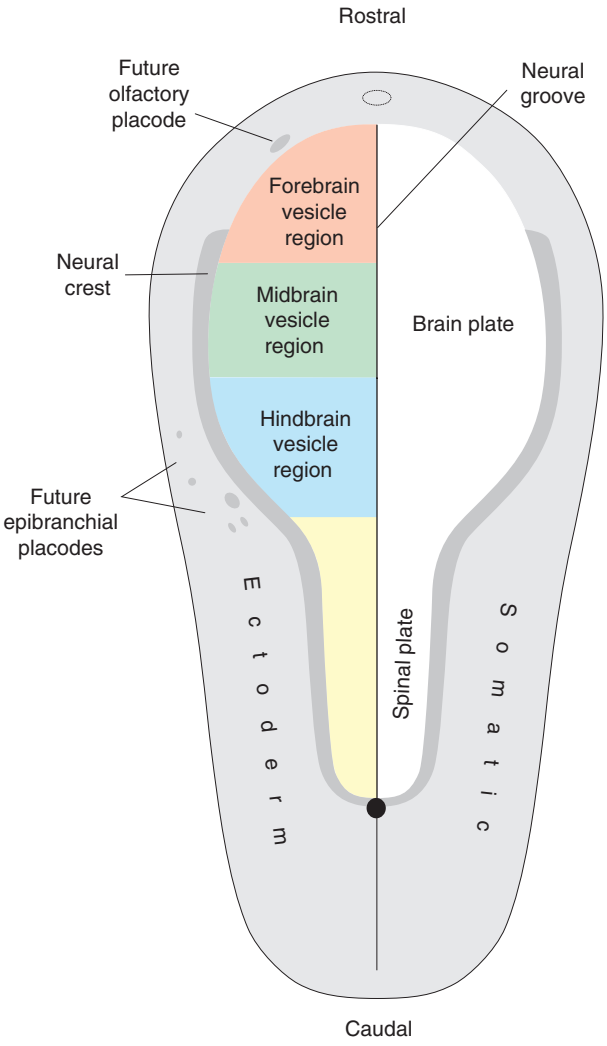


FIGURE 2.10 The neural plate is a spoon-shaped region of ectoderm (neural ectoderm) forming the CNS; surrounding it is somatic ectoderm. The neural plate is polarized (wider rostrally than caudally), bilaterally symmetrical (divided by the midline neural groove), and regionalized (brain plate rostrally, spinal plate caudally). The neural crest is a thin zone between neural and somatic ectoderm, and a series of placodes develops as "islands" within the somatic ectoderm. The neural crest and placodes generate PNS neurons. The approximate location of the future primary brain vesicles (Fig. 2.12A) in the neural plate is shown in color on the left. The same color scheme is used in Figs. 2.11, 2.12, and 2.14. *Refer to Swanson (1992).*

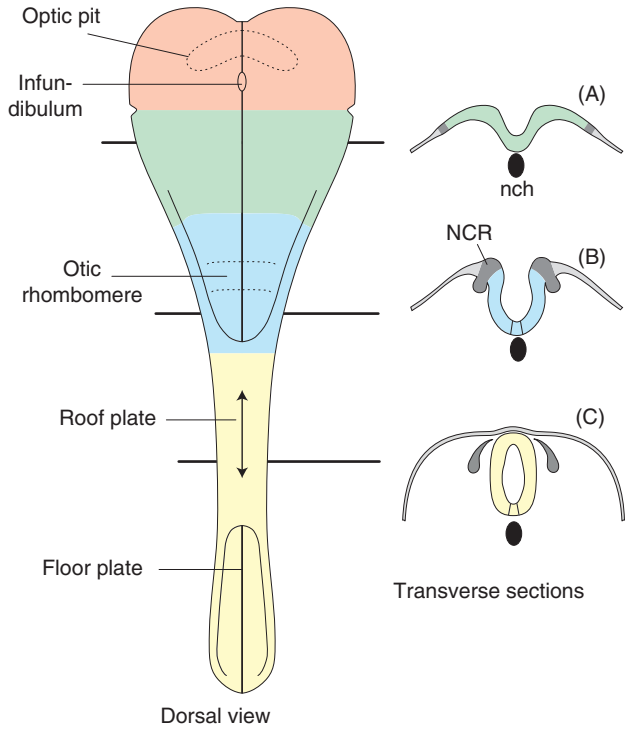


FIGURE 2.11 Optic pits, infundibulum, and otic rhombomere (dorsal view on left) are the earliest clear structural differentiations of the neural plate, other than the neural groove. The neural tube forms by neuroectoderm invagination (transverse sections A and B), followed by fusion of the lateral edges of the neural plate (roughly in the neck region of humans), and proceeds both rostrally and caudally (double arrows in roof plate). Note how the neural crest (NCR) pinches off in the process. Also observe notochord (nch) position ventral to neural groove. *Refer to Swanson (1992).*

Further Regionalization Occurs in the Neural Tube

As neurulation progresses, the neural plate becomes U-shaped as the two halves (*neural folds*) become vertically oriented (Fig. 2.11, right). Then the dorsal tips of the folds fuse, forming an open tube—and finally the tube's ends (*neuropores*) also fuse, producing a completely closed neural tube with a one-cell-thick wall (*neuroepithelium*).

Marcello Malpighi, the great seventeenth-century founder of histology, who also discovered the capillary network between arteries and veins postulated by William Harvey in 1628, recognized that the early chick neural tube displays three rostrocaudally arranged swellings now called primary brain vesicles. They include the *forebrain* (prosencephalic) vesicle, with the optic stalks and infundibulum; the *midbrain* (mesencephalic) vesicle; and the *hindbrain* (rhombencephalic) vesicle, with the otic rhombomere (Fig. 2.12A).

These vesicles are the fundamental structural or regional brain divisions. Transitory rhombomeres are the most characteristic primary hindbrain vesicle feature at this stage, and they develop in association with the pharyngeal pouches (Chapter 13). As embryogenesis continues, the forebrain vesicle divides into *endbrain* (telencephalic) and *interbrain* (diencephalic) vesicles, whereas the hindbrain vesicle becomes the *rhombicbrain*, which differentiates vaguely into two regions: rostrally the (secondary) *hindbrain* (epencephalon) and caudally the *medulla* (metencephalon); see Figure 2.12B. These divisions transform the "three primary vesicle stage" into

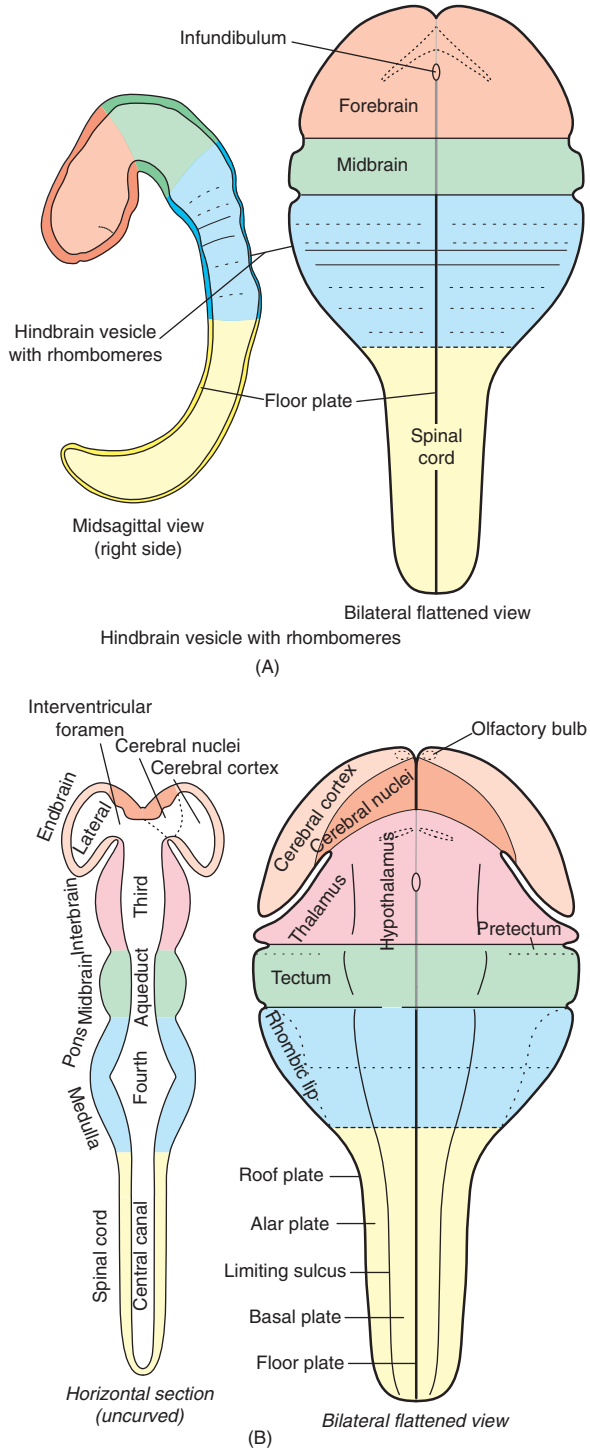


FIGURE 2.12 Formation and regionalization of the neural tube. (A) The early neural tube brain region develops three swellings: forebrain, midbrain, and hindbrain vesicles. The hindbrain vesicle then differentiates a series of transverse swellings called rhombomeres. (B) As differentiation continues, the forebrain vesicle displays right and left endbrain (cerebral hemisphere) vesicles and a medial interbrain vesicle, and the hindbrain vesicle shows vague pontine and medullary regions. This is the five-vesicle stage of neural tube transverse regionalization. Then longitudinal, dorsoventral, regionalization begins. The endbrain vesicle divides into cerebral cortex (including olfactory bulb) and cerebral nuclei (basal ganglia), the interbrain vesicle divides into thalamus and hypothalamus, the midbrain vesicle divides into tectum and tegmentum, the hindbrain vesicle becomes known as the rhombicbrain and divides into rhombic lip, alar plate, and basal plate, and the spinal cord divides into alar and basal plates. Whether the pretectal region (sometimes called synencephalon) is part of interbrain or midbrain is controversial. At this developmental stage major components of the adult ventricular system are seen in the neural tube lumen. *Refer to Swanson (1992) and Alvarez-Bolado and Swanson (1996).*

the "five secondary vesicle stage," with the hindbrain then producing a dorsal *cerebellum* and a ventral *pons*.

The neural tube lumen becomes the adult CNS *ventricular system* (Fig. 2.12B, left), and its adult shape conforms to extensive differential regionalization of the neural tube wall. Each endbrain vesicle contains a *lateral ventricle*, which communicates through an *interventricular foramen* with the *third ventricle* in the interbrain vesicle's center. The third ventricle continues into the midbrain's *cerebral aqueduct*, which becomes the rhombicbrain's *fourth ventricle* and then the spinal cord's *central canal*. In older embryos and adults, the ventricular system contains *cerebrospinal fluid* (CSF), much of which is elaborated by specialized, highly vascular regions of *choroid plexus* in the roof of the lateral, third, and fourth ventricles.

Migrating Neurons form the Mantle Layer's Gray Matter

In the early five secondary vesicle neural tube, cells divide repeatedly, although the neural tube remains one cell thick, a *pseudostratified epithelium* of stem cells for neurons and glia. Shortly thereafter many of these cells begin a terminal differentiation into young neurons migrating from the luminal proliferation zone to form a new, more superficial *mantle layer* (Chapter 16). In some CNS regions, mantle layer neurons segregate into layers parallel to the surface, whereas in other regions, neurons cluster in nuclei, relatively uniform neuron populations (usually multiple types) that are structurally distinct from surrounding nuclear or layered regions.

Mantle layer formation leads to further CNS regionalization (Fig. 2.12B). In rhombicbrain and spinal cord, it emerges because motoneurons are generated earliest and ventrally (corresponding to medial neural plate regions). This correlates with observations that gross, relatively uncoordinated embryonic motor behavior starts before reflex pathways become functional—and implying that such behavior is generated endogenously in the CNS itself (Hamburger, 1973).

Ventral mantle layer formation accompanies the transient appearance of a longitudinal groove (*limiting sulcus*) on the neural tube's inner surface. The leading nineteenth-century Swiss embryologist Wilhelm His noted that the limiting sulcus divides much of the neural tube into dorsal or *alar plate* and ventral or *basal plate*, with predominantly sensory and motor functions, respectively (Fig. 2.13). This observation complemented the earlier fundamental discovery of François Magendie that spinal sensory and motor fibers are completely segregated in spinal roots: sensory axons enter through dorsal roots, whereas motor axons leave through ventral roots. It is now clear that alar and basal plates are not

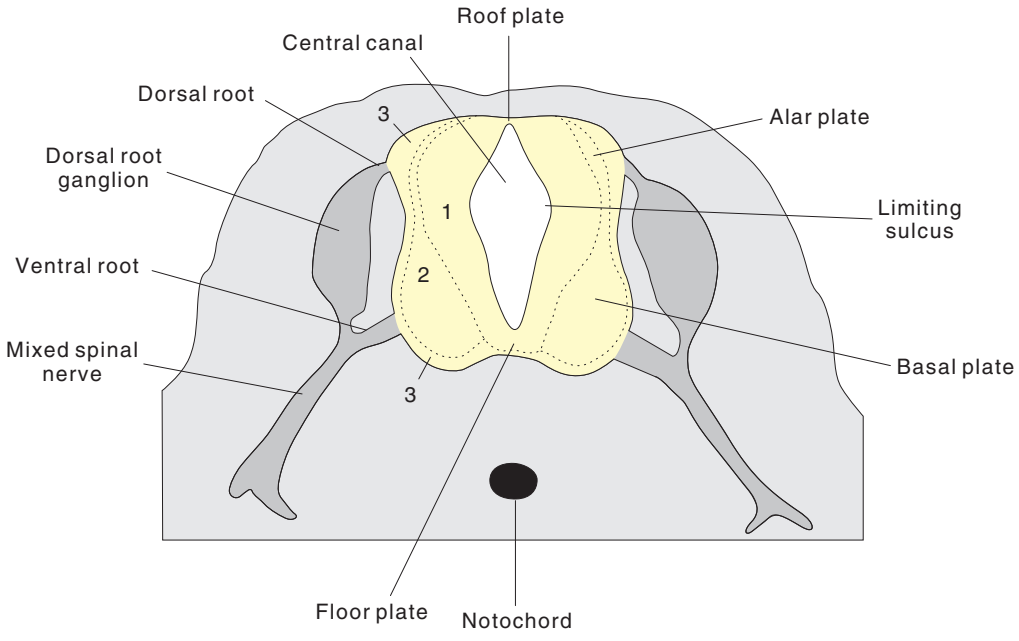


FIGURE 2.13 The early spinal cord and rhombicbrain are divided into dorsal (alar) and ventral (basal) plates by the limiting sulcus. This morphology reflects earlier ventral differentiation of the mantle layer (2), accompanied by earlier ventral thinning of the neuroepithelial or ventricular layer (1) of the neural tube, which remains as the adult ependymal lining of the ventricular system. The mantle layer develops into adult gray matter. This schematic drawing of a transverse spinal cord histological section also shows dorsal (sensory) and ventral (motor) spinal cord roots, dorsal root ganglia containing sensory neurons derived from the neural crest, and mixed (sensory and motor) spinal nerves distal to the ganglia. The peripheral area (3) is called the marginal zone and develops into the spinal cord white matter or funiculi containing ascending and descending axonal fiber tracts.

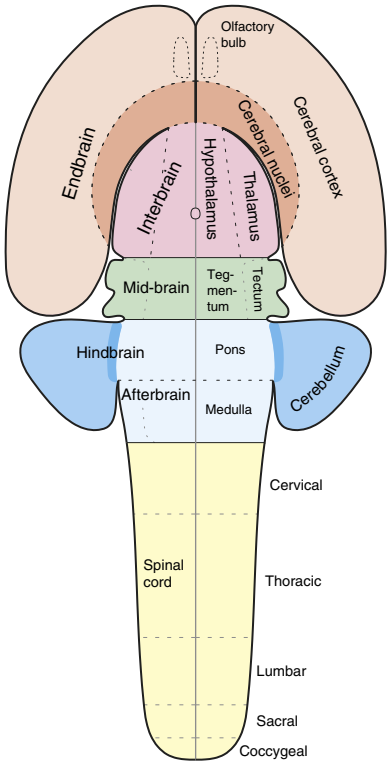


FIGURE 2.14 Major divisions of the adult mammalian CNS are derived from neural plate and neural tube regionalization illustrated in Figs. 2.10–2.12. *Modified from Swanson (1992).*

purely sensory or motor because each contains projection interneurons. Nevertheless, it is helpful to view the rhombicbrain and spinal cord as having three longitudinal zones: sensory, integrative (reticular formation), and motor. Regionalization of midbrain and forebrain does not fit this scheme neatly and is relatively poorly understood conceptually.

Dorsal regions of the rhombicbrain alar plate form a unique structure, the *rhombic lip*. In the rostral half it generates cerebellar granule cells, whereas more caudally it produces neuron populations like the precerebellar and vestibulocochlear nuclei. Many rhombic lip neuron populations are interesting because they migrate tangentially, parallel to the neural tube's surface to reach their final destinations, instead of radially (transversely) like most CNS neurons (Chapter 16). This differentiation continues until the adult CNS configuration is achieved (Figs. 2.14 and 2.15). The most obvious late-developing structures are the cerebral cortex and cerebellar cortex.

Summary

The vertebrate CNS develops from a sheet of cells called the neural plate that invaginates to form the neural tube. The tube's rostral end differentiates a series of vesicles that constitute the major brain regions, and the caudal end forms the simpler spinal cord. Most PNS neurons differentiate from the neural crest, with the rest arising from nearby somatic ectodermal placodes.

THE BASIC PLAN OF NERVOUS SYSTEM CONNECTIVITY

Functional Systems Consist of Interconnected Gray Matter Regions

The nervous system's wiring diagram can be described in terms of the neuron types in each of its distinct gray matter regions and their stereotyped pattern of axonal connections to cell types both locally (within the region) and in other gray matter regions or other tissues (like muscle or gland). A long-term goal of systems neuroscience is to provide a global wiring diagram for the nervous system that systematically accounts for its various functional subsystems—analogous to the circulatory system model provided by Harvey. Little work has been done on this synthetic problem, although interest is accelerating with the development of online neuroinformatics workbenches for connectional information (Bota & Swanson, 2010) and with plans for complete tables of nervous system connections—connectomes (Sporns, Tonono, & Kötter, 2005).

The high-level model of nervous system information processing shown in Figure 2.16 synthesizes basic neurobiological concepts pioneered by Cajal and Sherrington with basic cybernetic principles pioneered by Norbert Wiener (1948) and John von Neumann (1958). In essence, the model postulates that behavior is determined by CNS motor system output and that this output is a function of three inputs: sensory system (reflexive), cognitive system (voluntary), and intrinsic behavioral state system. The relative importance of each input in controlling motor output (behavior) varies qualitatively in different species and quantitatively in different individuals. Note that behavior elicits sensory feedback from the external and internal environments that helps determine future motor activity and thus behavior. Each component is now considered further without trying to place all nervous system parts within the global model.

Motor Systems are Organized Hierarchically

There are three different motor systems: skeletal, autonomic, and neuroendocrine. The first controls striated muscles responsible for voluntary behavior; the second controls smooth and cardiac muscle, and many glands; and the third controls pituitary gland hormone secretion. The skeletal motor system is understood best and thus serves as a prototype for examining basic organizing principles presumably similar for all three.

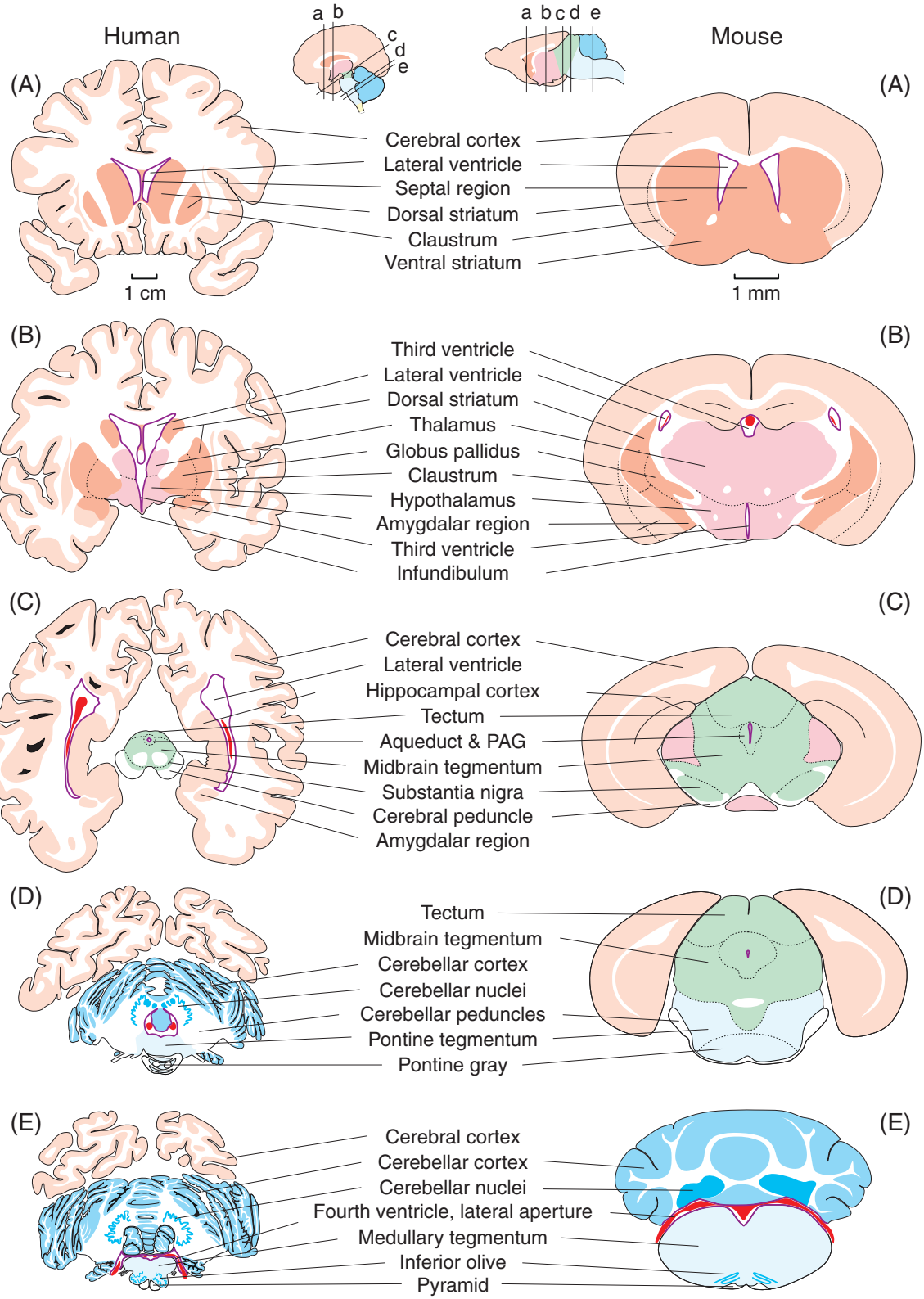


FIGURE 2.15 Mini atlases to compare major adult brain regions in humans and mice. The brains are cut approximately transversely to the CNS longitudinal axis and illustrate five major levels, arranged from rostral to caudal: a, endbrain; b, interbrain; c, midbrain; d, pons; and e, medulla. The color scheme follows that in Figs. 2.10–2.12 and 2.14, with the choroid plexus of the lateral, third, and fourth ventricles shown in red. *Adapted from Nieuwenhuys, Voogd, and Van Huijzen (1988) and Sidman, Angevine, and Webster (1971).*

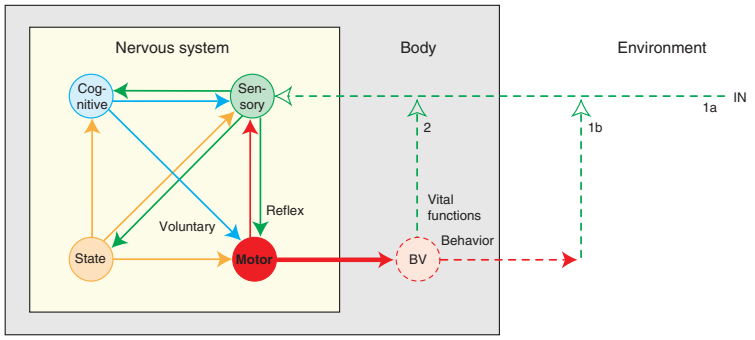


FIGURE 2.16 A model of the nervous system's basic wiring diagram. The model of information flow through the nervous system (yellow box) postulates that behavior and vital body functions are controlled by the motor system (red), which is influenced by three classes of neural input: sensory (green), intrinsic behavioral state (gold), and cognitive (blue). Sensory inputs lead directly to reflex responses, cognitive inputs mediate voluntary responses, and intrinsic inputs act as control signals to regulate behavioral state. Motor system outputs produce behaviors whose consequences are monitored by sensory feedback (1b, 2). Sensory feedback may be used by the cognitive system for perception and by the intrinsic system to generate affect (e.g., positive and negative reinforcement/pleasure and pain). The cognitive, sensory, and intrinsic systems are all strongly interconnected in the indicated pattern. *Refer to Swanson (2012).*

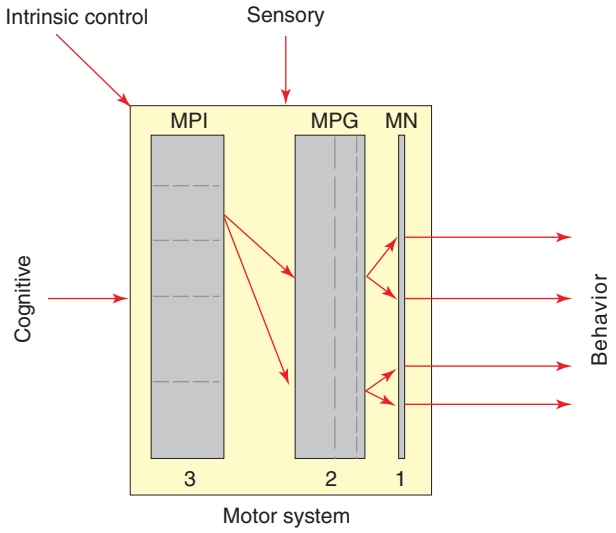


FIGURE 2.17 Hierarchical organization of the skeletal motor system. At the simplest level (1), motoneuron pools (MN) innervate individual muscles generating individual components of behavior. At the next higher level (2), additional interconnected interneuron pools, called motor pattern generators (MPG), innervate specific motoneuron pool sets. At the highest level (3), additional interconnected interneuron pools, called motor pattern initiators (MPI), innervate specific MPG sets. MPIs can activate complex, stereotyped behaviors when activated (or inhibited) by specific patterns of sensory, intrinsic, and/or cognitive inputs. Note that MPGs and MPIs themselves may be organized hierarchically (dashed lines) and that sensory, intrinsic, and cognitive inputs may go directly to any level of the motor system hierarchy. *Refer to Swanson (2012).*

The skeletal motor system is arranged hierarchically (Fig. 2.17), the lowest level consisting of brainstem-spinal cord α-motoneurons whose axons synapse directly on striated muscle fibers. The next higher level consists of motor pattern generators (MPGs), and the highest level has motor pattern initiators (MPIs) that "recognize" or alter their output in response to specific input patterns and project to unique sets of MPGs. Ethologists refer to MPIs as "innate releasing mechanisms." One reason central neural circuitry is so complex is that each of the three input types (sensory, intrinsic, cognitive) may go directly to each general level of the motor system hierarchy.

The MPGs and MPIs themselves are hierarchically arranged. This organization is particularly easy to see conceptually for the MPGs subserving locomotor behavior. In the spinal cord, simple MPGs coordinate the reciprocal innervation of muscle pair antagonists across individual joints, more complex MPGs coordinate activity in the set of simpler MPGs for all the joints in a limb, and still more complex MPGs coordinate activity in MPGs for all four limbs. At the next higher level there is a brain hierarchy of MPIs for locomotion that is activated by specific input patterns and projects to the spinal locomotor pattern generator network.

Multiple Sensory Systems Function in Parallel

A set of sensory systems provides information to the CNS from various receptor types, and all the systems can function simultaneously. Cajal noted that unimodal sensory pathways generally branch with some information going directly to the motor system and some going to the cerebral cortex for sensation and perception. The former typically evokes reflex behavior, and the latter potentially reaches consciousness and plays an important role in cognition.

Several general features characterize the sensory system (Section IV covers subsystems in detail). First, the CNS receives a wide range of information about the external environment and about the body's internal state. Thus, sensory receptors lie near the body's surface (e.g., touch and olfactory receptors), deep within the body (e.g., aortic stretch receptors), and even within

the brain itself (e.g., hypothalamic insulin receptors). Second, each of the three motor systems receives a broad range of sensory inputs. Third, the range of sensory modalities is remarkably similar (though not identical) across vertebrate classes, and information about specific modalities enters the CNS through homologous cranial and spinal nerves in all vertebrates. And fourth, the number of synapses between sensory receptor and cerebral cortex varies in different systems. There is one synapse in the olfactory system and at least four in the visual system.

The Cognitive System Generates Anticipatory Behavior

It is very likely that the cerebral cortex—along with its cerebral nuclei (basal ganglia)—is the most important, if not sole, part of the cognitive system and that the cerebral cortex is responsible for planning, prioritizing, initiating, and evaluating the consequences of voluntary behavior (Section VII). The fundamental nature of voluntary behavior is obviously a difficult problem to address, but one useful approach is simply to compare it with reflexive behavior. Interestingly, most if not all behaviors mediated by skeletal muscle can be initiated either reflexively or voluntarily, as Descartes pointed out long ago. What seems to distinguish reflexive and voluntary behaviors most clearly is that the former involves a stereotyped response to a defined stimulus, whereas the latter is anticipatory, with a duration and content impossible to predict with anywhere near the same degree of certainty.

Intrinsic Systems Control Behavioral State

The CNS generates considerable endogenous activity (action potential patterns); it is definitely not just a passive system waiting to respond to sensory input, as the behaviorist approach a century ago assumed. All CNS parts apparently have a basal activity level that can be either increased or decreased. In many cases, it is still not established whether particular neuron types generate intrinsic activity patterns. It is clear, however, that motoneurons and related MPGs do generate intrinsic activity; as already noted, the embryonic spinal cord produces motor output before sensory circuits develop. Thus, in addition to the three extrinsic input types to the motor system illustrated in Figure 2.16, intrinsic activity within the motor system itself can produce behavior that is neither reflexive nor voluntary.

Certain CNS regions generate intrinsic rhythmic activity patterns. The most important rhythmic behavioral pattern is the sleep–wake cycle that is entrained to the day-night cycle by an endogenous circadian clock, the hypothalamic suprachiasmatic nucleus (Chapters 39 and 40). The sleep–wake cycle is profoundly important because during sleep the body is maintained entirely by ongoing intrinsic and reflexive systems controlling behaviors like respiration and sustained sphincter contractions. In contrast, voluntary mechanisms dominate in wakefulness though reflexive and intrinsic mechanisms are also vitally important then.

Behavioral state control is thus a fundamental intrinsic brain activity. Another aspect of behavioral state—arousal—is especially important during wakefulness. Arousal level generally is correlated with an animal's motivational state or drive level (Chapter 41). The neural system mediating drive is not fully elucidated but is critically dependent on the hypothalamus, and attainment of specific goal objects (foraging behavior) depends on the cognitive system. Arousal and drive may be controlled by subcortical systems but behavior's actual direction and prioritization mainly is determined cortically.

The full identity of neural systems elaborating pleasure and pain is one of neuroscience's deep mysteries. Many regard pleasure and pain as conscious expressions of positive and negative reinforcement, influencing how likely a particular voluntary behavior will be repeated or avoided in the future. Here, reinforcement depends on sensory feedback about a particular behavior's consequences (Fig. 2.16), and one suggestion is that pleasurable and painful sensations, like those associated with drive, are elaborated subcortically within intrinsic control systems. According to this view, thinking or cognition arises in cerebral cortex, whereas feeling or affect arises subcortically. It is also possible that all aspects of consciousness (thinking and feeling) arise only from cortical neural activity (Chapter 51).

How Pharmacological and Genetic Networks Relate to Functional Systems

Specific neurotransmitter systems have been incorporated into models of CNS function since the 1950s. Two examples are cholinergic and noradrenergic systems, defined as the total sets of CNS neurons releasing acetylcholine or noradrenalin, respectively, as a neurotransmitter (Chapter 5). In general, these systems are not obviously correlated with traditional CNS functional systems or major topographic parts; typically they are not restricted to one functional system or one major CNS division, though some exceptions may exist. Thus, neurotransmitter systems are not functional systems in the traditional sense. However, they are conceptually or operationally important in helping define networks or functional systems influenced by particular drug actions. For example, administering centrally acting acetylcholine receptor agonists influences synapses in a variety of traditional *functional systems*, and the set of these functional systems could be defined as a *pharmacological system* with a specific set of behavioral and other

responses. If a drug is targeted for therapeutic reasons to a specific neural system (e.g., a cholinergic agonist targeted to the cerebral cholinergic system in Alzheimer's disease; Chapter 43), it will also act on other functional systems with appropriate cholinergic receptors (e.g., in the thalamus and lower brainstem). Responses in these other systems produce "side effects" that may be good or bad.

Likewise, any gene product's distribution pattern can also be used to define a chemical, molecular, or neural *gene expression system*. For example, a system could be defined in terms of all neurons expressing the calbindin or μ-opioid receptor gene, and expression of the corresponding gene might be prevented or altered in experimental knockout mice or natural mutations in genetic diseases. These alterations may produce an obvious and stereotyped phenotype or syndrome, but in most instances the gene normally is expressed in multiple functional systems and has complex (even if subtle) physiological and behavioral effects.

Finally, it is important to remember that a genetic program constructs the nervous system's basic *macrocircuitry* during embryogenesis. Determining the correspondence between gene expression networks and neural networks may be the ultimate achievement of systems neuroscience. The nervous system's *microcircuitry*—quantitative aspects of synapse number and strength associated with individual neurons—may be sculpted by experience throughout life.

Summary

There is no simple relationship between the CNS's topographic or regional differentiation and its functional organization. So it is mistaken to assume *a priori* that CNS information simply is processed hierarchically with the spinal cord at the lowest level and the cerebral cortex at the top. An alternative view is that the CNS displays a network rather than hierarchical organization scheme—a circuit where the motor system is driven by sensory, cognitive, and intrinsic behavioral state inputs, and future motor activity is determined partly by sensory feedback about the initial behavior's consequences.

Two major features complicate this simple network model. First, the motor system itself is organized hierarchically, whereas the sensory system transmits multiple modalities in parallel, and this sensory information can reach directly each level of the motor system hierarchy. And second, sensory information also reaches the intrinsic and cognitive systems. In fact, all three input systems are interconnected in a genetically determined pattern. The basic plan of neural circuit architecture must be understood on its own terms, not through simple preconceived ideas or superficial analogies with computers, the Internet, irrigation systems, or complicated robots. How traditional CNS functional systems relate to pharmacological systems and genetic networks remains to be determined.

OVERVIEW OF THE ADULT MAMMALIAN NERVOUS SYSTEM

This section reviews structural neuroscience methods used to achieve our current—still very incomplete—understanding of nervous system architectural principles and introduces the major nervous system components. Long experience teaches that nothing approaches actual dissection for gaining an appreciation of overall brain structure.

A Brief History of Structural Neuroscience Methods

The human brain's macroscopic structure was observed by early Greek physician-philosophers and thoroughly understood by the early 1800s. However, its circuitry's astounding complexity was not appreciated until microscopy effectively identified individual pathways (axon bundles) and neuronal regions (distinguishable neuronal cell body aggregates) toward the end of the 1800s—applying to thin CNS tissue sections neurohistological reagents and reactions developed by the textile and photographic industries (Swanson, 2000b).

Perhaps the single most enduring contribution of nineteenth-century neurohistology was Camillo Golgi's 1873 silver impregnation method. The full morphology of individual neurons was visible for the first time (Fig. 2.18; Boxes 2.1 and 2.2)—their dendritic tree, cell body shape, axon and all its collaterals, and points of presumed functional contact with other cells, which Sherrington named "synapses" in 1897. The Golgi method involves impregnating brain tissue alternately in potassium dichromate and silver nitrate solutions over periods of weeks to years, and mysteriously (the reaction's chemistry remains elusive) about 1% of the neurons are filled (apparently randomly) with a dense precipitate. The Golgi stain thus reveals more by staining less. Today, selective labeling of individual neurons also is achieved by injecting markers directly into living neurons with micropipettes, allowing simultaneous electrophysiological and cytoplasm analysis.

Unfortunately, Golgi's method provided little information about longer CNS connections: axonal projections between nonadjacent regions. This was approached with methods selectively staining fibers degenerating from pathological or experimental lesions. Augustus Waller showed (1850) that nerve transection causes the nerve's distal segment to degenerate (Wallerian, anterograde degeneration), inspiring his proposal that the cell body is

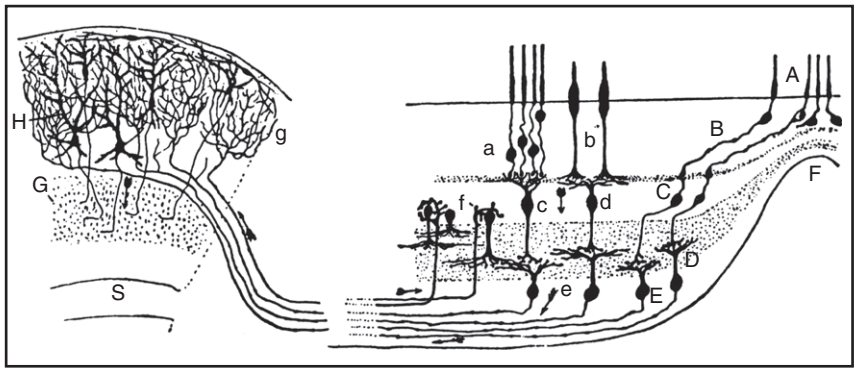


FIGURE 2.18 Cajal's (1909–1911) neural architecture drawing based on the Golgi method. It shows the organization of four major retinal neuron types (right) and connections between two gray matter regions: retina to optic tectum (superior colliculus; left). Applying the neuron doctrine and functional polarity rule (Boxes 2.1 and 2.2) to the entire vertebrate nervous system by Cajal and many other researchers a century ago led to the "classical" way neuron types defining each gray matter region have been described structurally ever since, a view beautifully illustrated here. Three major neuron types tend to populate specific retinal layers: photoreceptors (subtypes a, b, A, B), bipolar cells (subtypes c, d), and ganglion cells (subtypes e, D, E). A specific neuron type's cardinal feature is its axon's distribution—the neuron's function in terms of output. Photoreceptors detect light and their axon innervates bipolar cells. The latter in turn innervate ganglion cells whose axon projects through the optic nerve to the tectum. Photoreceptors are classical sensory neurons (Fig. 2.3), bipolar cells are local interneurons, and ganglion cells are projection interneurons. Also note a second retinal local interneuron class, amacrine cells (f). Cajal pointed out that retinal neuronal cell bodies aggregate in three layers with synaptic neuropil zones in between, and illustrated a clear structural gradient—reflecting a foveal region (F) with greater visual acuity because of multiple structural features, some of which are obvious in the drawing. The power of Cajal's functional polarity theory is evident: he drew arrows to indicate the presumed normal direction of information flow through the network, based on the sequential arrangement of dendrites and axons associated with each neuron type.

the nerve cell's "trophic center," which the axon depends on for survival. Thirty years later Bernard von Gudden showed that Wallerian degeneration may be accompanied by pathological changes in the cell body when its axon is cut, suggesting retrograde transport of "trophic factors" from axon to cell body.

Marchi and Algeri developed the first method to stain selectively central pathways (1885), revealing degenerating myelin sheaths of severed axons as black particles on a light background—effectively isolating degenerating from healthy sheaths in tissue sections. Anatomists could produce discrete CNS lesions in experimental animals and after several weeks trace the course of neural pathways arising in the lesioned site. The method was severely limited, however, because unmyelinated or thinly myelinated axons, and terminal fields, were unlabeled, and there were many "false-positive" results from transecting fibers simply passing through the lesion (the "fiber-of-passage" problem common to many experimental methods).

By the 1950s selective silver impregnation and degeneration methods were combined by Walle J.H. Nauta and colleagues to stain unmyelinated axons and their terminal fields. The CNS was remapped at finer resolution with these methods, which still suffered from the fiber-of-passage problem (false-positive results) and were not nearly as sensitive (false-negative results) as the next generation of methods developed around 1970. Instead of relying on lesion-induced pathology the latter (current) are based on a combination of (1) physiological mechanisms (anterograde and retrograde intraaxonal marker transport) in healthy neurons and (2) histochemical detection of antibodies and complementary nucleic acid strands.

Also in the 1950s, the electron microscope opened a whole world of ultrastructure previously only guessed at. It provided the first glimpses of synapse structure (with a typical cleft only about 20 nm wide, far below the 1 mm resolution of light microscopes, and presynaptic vesicles), myelin sheath organization, and many cellular organelles. It also allowed biologists to examine in detail the biosynthetic apparatus residing within each cell.

These methods provide far more detail about CNS connectivity patterns or circuit organization than ever before in many species. As a result, comparative neuroanatomy has flourished and forms a solid structural foundation on which contemporary physiological and behavioral studies are based (Luo, Callaway, & Svoboda, 2008).

The PNS has Sensorimotor and Autonomic Divisions

Overall, the nervous system is divided into CNS (brain and spinal cord) and PNS (nerves, ganglia, and related plexuses). However, this CNS–PNS distinction is just a gross anatomical convenience that ignores circuit organization because nerves contain axons from neuronal cell bodies in both CNS and peripheral ganglia.

Functionally, the PNS has (1) a sensory ganglion component with accompanying dorsal and ventral roots and functionally mixed nerves, (2) the autonomic nervous system's (ANS) motor ganglia and communicating roots, and (3) plexuses (literally "braids"). The sensory part has dorsal root ganglia sensory neurons with one extension (embryologically and phylogenetically an axon) entering the CNS through dorsal roots and the other extension (embryologically and phylogenetically a dendrite) traversing peripheral nerves to sites throughout the body. However, peripheral nerves also contain axons of skeletal motoneurons with cell bodies in spinal cord and brainstem; for spinal nerves the axon's initial part traverses a ventral root. Thus, most peripheral nerves carry *afferent* ("sensory") information toward the CNS and *efferent* (motor) information toward the body (Fig. 2.13). Sensory neurons carry afferent information from receptors in skin, skeletal muscles, tendons, joints, blood vessels, and deep viscera. The autonomic nervous system has a network of efferent pathways, ganglia, and nerve nets (plexuses) controlling gut peristalsis, glandular secretions, blood vessel diameter, and other visceral functions—and their output is modulated by both somatic and visceral afferents. Thus, a typical peripheral nerve carries a mixture of afferents and efferents innervating body wall and viscera.

Somatic afferents distribute near the body surface in a pattern reflecting the body wall's segmental origins: each spinal nerve innervates a narrow mediolateral band of skin called a *dermatome* (Fig. 25.9), although adjacent nerves innervate overlapping territories (otherwise single nerve interruption would produce complete sensation loss in a band of skin, which is not the case). The segmental dermatome pattern is obvious in the torso where very little differential body wall growth occurs, whereas limb dermatomes are distorted because they form before the limbs grow out fully in the embryo.

Peripheral nerves often ramify and join with nerves from other segments to form *nerve plexuses* that serve as crossroads and distribution centers for the nerves, allowing axons to reorganize into complex nerve bundles innervating body structures. Brachial and lumbosacral plexuses near the origins of the upper and lower limbs, respectively, are the largest examples of these enigmatic structures that also provide great mechanical strength to nerves passing through the shoulders and hips, which may undergo extreme rotation and stretching. The ANS also has a bewildering variety of *ganglionated plexuses* in the abdomen and pelvis, where nerves converge and redistribute axons to their target organs and where individual neurons and/or clumps of autonomic neurons are found.

The ANS has structural and functional *sympathetic* and *parasympathetic* divisions (Chapter 34). The two divisions function in a kind of push-pull relationship with each other. One or the other is never completely on or off. Instead, there are degrees of sympathetic and parasympathetic tone. During sleep, certain involuntary functions like digestion are accelerated. Glands participating in digestion are activated parasympathetically and sympathetic tone is correspondingly decreased. In contrast, Walter B. Cannon noted almost a century ago that during the "fight or flight" reaction characterizing defensive behavior, sympathetic tone is markedly enhanced and parasympathetic tone is reduced sharply. Sympathetic outflow is vastly amplified and coordinated through a set of ganglia and the adrenal medulla, so that sympathetic function occurs relatively synchronously throughout the body. In contrast, the parasympathetic system is relatively finely tuned.

The Cerebrospinal Trunk Generates Cranial and Spinal Nerves

From a more systematic perspective on nervous system organization, the spinal cord and brainstem (together the cerebrospinal trunk) generate a continuous series of spinal and cranial nerves, respectively. The human spinal cord, roughly as thick as an adult's little finger, is ultimately surrounded and protected by the vertebral column, whereas the cranial part of the skull protects the brain. In cross section, the spinal cord's two basic types of nervous tissue are obvious: *gray matter* and *white matter*. Gray matter forms an H-shaped region surrounding the central canal (the ventricular system's spinal segment) and consists mainly of neuronal cell bodies and neuropil. White matter surrounds gray matter in the spinal cord and consists mostly of axons collected into overlapping fiber bundles. Many axons have a myelin sheath, a uniquely vertebrate feature allowing rapid nerve impulse conduction (Chapter 4) and giving white matter its pale appearance.

The spinal cord looks segmented because bilateral pairs of dorsal and ventral roots emerge regularly along its length. These pairs form five sets: cervical (in the neck above the rib cage), thoracic (associated with the rib cage), lumbar (near the abdomen), sacral (near the pelvis), and coccygeal (associated with tail vertebrae). In humans there are typically 31 spinal nerve pairs (8 cervical, 12 thoracic, 5 lumbar, 5 sacral, and 1 coccygeal) that are named according to the intervertebral foramen they pass through. This enumeration varies between species.

Based on human brain macroscopic dissection, Samuel Thomas von Sömmerring in 1778 recognized a sequence of 12 cranial nerve pairs, and his basic classification scheme remains traditional for vertebrates in general, although it is problematic in terms of completeness (e.g., not including the nasal cavity's terminal nerve) and nonconformance with contemporary fate maps of cranial nerve nucleus development (e.g., motoneurons

for nerve VII are generated rostral to those for nerve VI). In any event, cranial nerves are more heterogeneous functionally than spinal nerves, and indeed most cranial nerve pairs have distinct compositions in terms of functional fiber types.

In humans (Fig. 2.19), seven cranial nerves transmit information about the so-called special senses associated with the head: olfaction (I, olfactory nerve—purely sensory, arising in nasal olfactory epithelium), vision (II, optic nerve from retina), hearing and balance (VIII, vestibulocochlear nerve from inner ear), and taste (V, VII, IX, and X; parts of the trigeminal, facial, glossopharyngeal, and vagus nerves, respectively). Nerves III (oculomotor), IV (trochlear), and VI (abducens) primarily control conjugate eye movements, although the third nerve also mediates autonomic control of the pupillary light reflex and lens accommodation. Major parts of the trigeminal (V) nerve carry sensory axons from the face (a rostral extension of the spinal somatosensory system) and motor axons innervating the muscles of mastication (chewing). The intermediofacial (VII) nerve controls the muscles of facial expression (facial nerve) and also innervates the salivary and lacrimal glands (intermediate nerve, VII′)—its role in emotional expression is obvious. The glossopharyngeal (IX) nerve innervates the tongue and pharynx and mediates the swallowing reflex. The vagus ("wandering," X) nerve has an exceptionally complex and widespread innervation pattern, including laryngeal muscles producing speech, and the parasympathetic innervation of most thoracic and abdominal viscera. The accessory (XI) nerve innervates several muscles that stabilize the head and neck and the hypoglossal (XII) nerve innervates the tongue musculature.

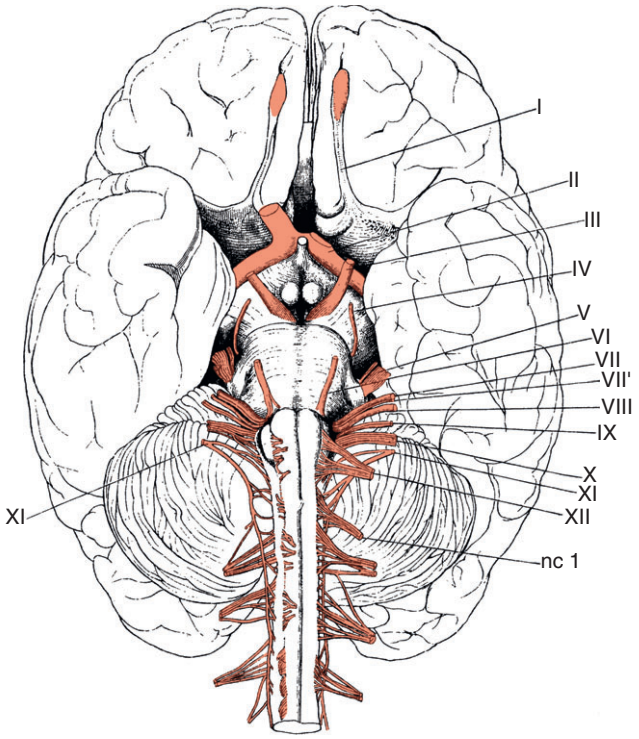


FIGURE 2.19 Origins of the cranial nerves viewed from the base of the adult human brain. The traditional numbering system dating back to the eighteenth century is indicated, but does not consider the terminal and accessory olfactory nerves. I, indicates the olfactory peduncle, with the olfactory bulb at the end shaded pink to indicate the termination site of the olfactory nerve; II, the optic tract leading from the optic chiasm to the thalamus and superior colliculus, with the stump of the right optic nerve indicated but the left optic nerve cut off at the chiasm; III, oculomotor nerve; IV, trochlear nerve, which courses around the side of the cerebral peduncle from the dorsal surface of the brainstem; V, trigeminal nerve; VI, abducens nerve; VII, facial nerve component of intermediofacial nerve; VII′, intermediate nerve component of intermediofacial nerve; VIII, vestibulocochlear nerve; IX, glossopharyngeal nerve; X, vagus nerve; XI, accessory nerve, with cranial and spinal roots; and XII, hypoglossal nerve; nc I, first cervical spinal nerve with the second through fourth also shown arising more caudally from the spinal cord. Cranial nerves exit through holes in the cranium whereas spinal nerves exit between the vertebrae (the first cranial nerve exits between the cranium and first vertebra). *Adapted from Henle (1871).*

Cerebral Hemispheres and Cerebellum are Divided into Cortex and Nuclei

Macroscopically the mammalian cerebrospinal trunk has two great expansions—the cerebral hemispheres and cerebellum—and both have an outer laminated cortex surrounding deep nonlaminated nuclei. The most extraordinary growth of the mammalian brain occurs in the endbrain or cerebral hemispheres (Fig. 2.20), which develop more or less as mirror images of one another and are separated in the dorsal midline by a deep interhemispheric (longitudinal) fissure. In humans, the sulcal pattern is, however, asymmetric and unique in each person, and there are functional asymmetries as well; for example, the speech centers typically are lateralized (Chapter 49). Hemisphere volume is restricted by skull capacity, so as the hemispheres grow during embryogenesis they develop folds (*gyri*) separated by invaginations (*sulci*, and when deeper, *fissures*). This corrugation allows cerebral (and cerebellar) cortex to have a larger surface area. The extent and pattern of folding vary stereotypically with species, although like any trait there are quantitative differences between individuals of a particular species. Two major grooves, the central sulcus and lateral (Sylvian) fissure, are used as anatomical landmarks in the human cerebrum. The central sulcus extends roughly vertically along the hemisphere's lateral surface where it approaches the horizontally oriented lateral fissure. Together they divide arbitrarily the lateral cerebral cortical surface into four lobes (*frontal*, *parietal*, *occipital*, and *temporal*), named for the overlying cranial bones. In addition, the *insular lobe* is folded completely inside the hemisphere, deep to the lateral fissure (actually about two-thirds of the folded cortical surface lies buried and unexposed to the outer hemisphere surface), and

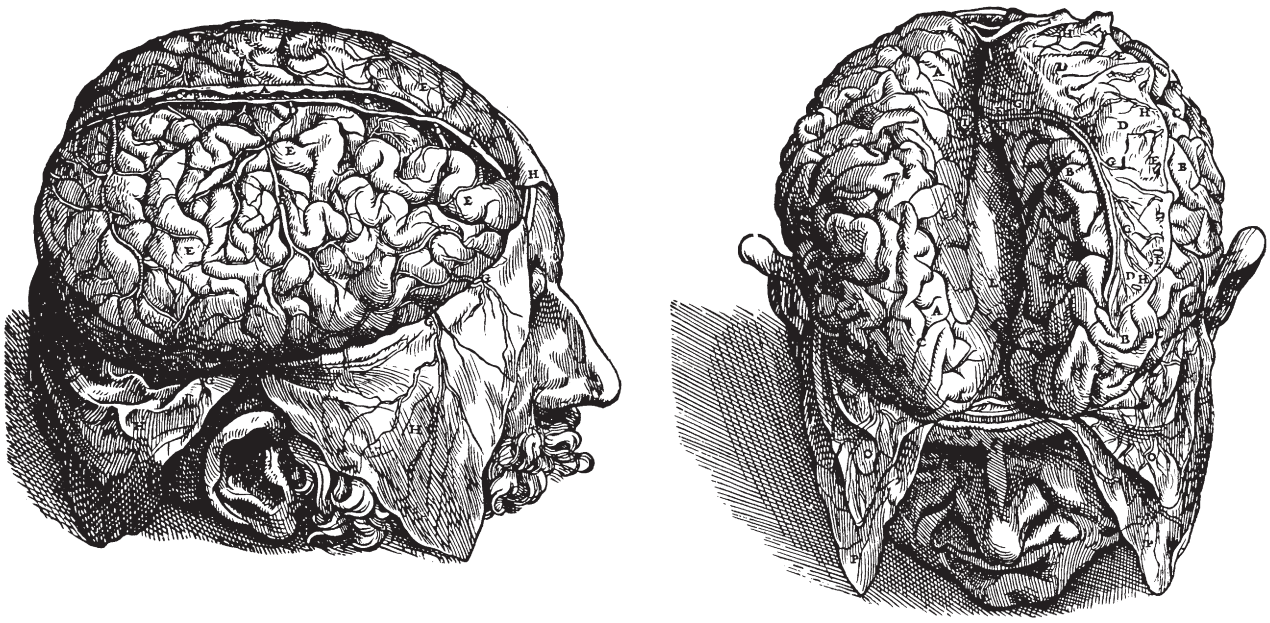


FIGURE 2.20 Surface features of the human cerebral cortex, which is thrown into gyri separated by sulci. In the drawing on the right, the right and left hemispheres have been pulled apart at the interhemispheric or longitudinal fissure to reveal a white matter tract, the corpus callosum (L), interconnecting the two hemispheres. The drawings are from perhaps the most important book in the history of medicine by Andreas Vesalius, *Fabric of the Human Body*, published in 1543. The drawings were probably executed by an artist from Titian's studio. See Singer (1952).

the *limbic lobe* forms the hemisphere's medial border along the interhemispheric fissure.

These lobes are only crude guides to the cerebrum's functional organization. Over the last 150 years, progressively better analysis has parceled the cortical mantle into a mosaic of roughly 50 to 100 areas with more or less distinct structural and functional characteristics. The most famous and enduring cortical *regionalization maps* were generated by Korbinian Brodmann a century ago (Fig. 2.21), although refinements and alternative interpretations abound. Nevertheless, cortical regionalization maps are fundamentally important guides for understanding CNS architecture. Just as one example, virtually the entire thalamus projects topographically on the cortical mantle, which in turn projects topographically on the entire cerebral nuclei (basal ganglia). Information from every sensory modality reaches the cerebral cortex and it in turn sends inputs to virtually the entire motor system.

Most cerebral cortical areas directly modulate activity on the opposite (contralateral) side of the body through descending pathways that cross the midline to reach motor system parts in the contralateral CNS. Furthermore, axon bundles called *commissures* connect cerebral cortical areas of one hemisphere with the same or related areas of the opposite hemisphere—and different areas in the same hemisphere are interconnected through complex *association pathways*. Thus, commissural and association pathways allow comparison and integration of information between cortical areas within and between the cerebral hemispheres. Descending, commissural, and association connections tend to arise from different sets of pyramidal neurons, which helps explain the highly laminated organization of the cerebral cortex: descending connections arise mostly from the deep layers 5 and 6, commissural connections from layer 3, and association connections from layer 2.

Communication between hemispheres is eliminated by commissurotomy, the surgical division of all cerebral commissures (including the hippocampal and anterior commissures). This procedure sometimes is used to treat otherwise intractable epilepsy cases, preventing spread of severe epileptic activity from one hemisphere to the other. Incredibly, commissurotomy patients function very well most of the time, and behavioral studies on such "split brain" patients have yielded remarkable information about cerebral cortical organization (Gazzaniga, 2005). Axon bundles (tracts or pathways) connecting very different structures on the two sides of the CNS usually are called *decussations* to distinguish them from commissures.

The Nervous System is Protected by Membranous Coverings

The CNS is completely surrounded by three concentric connective tissue membranes: *pia*, *arachnoid*, and *dura*. The pia ("faithful") is a very thin layer that adheres

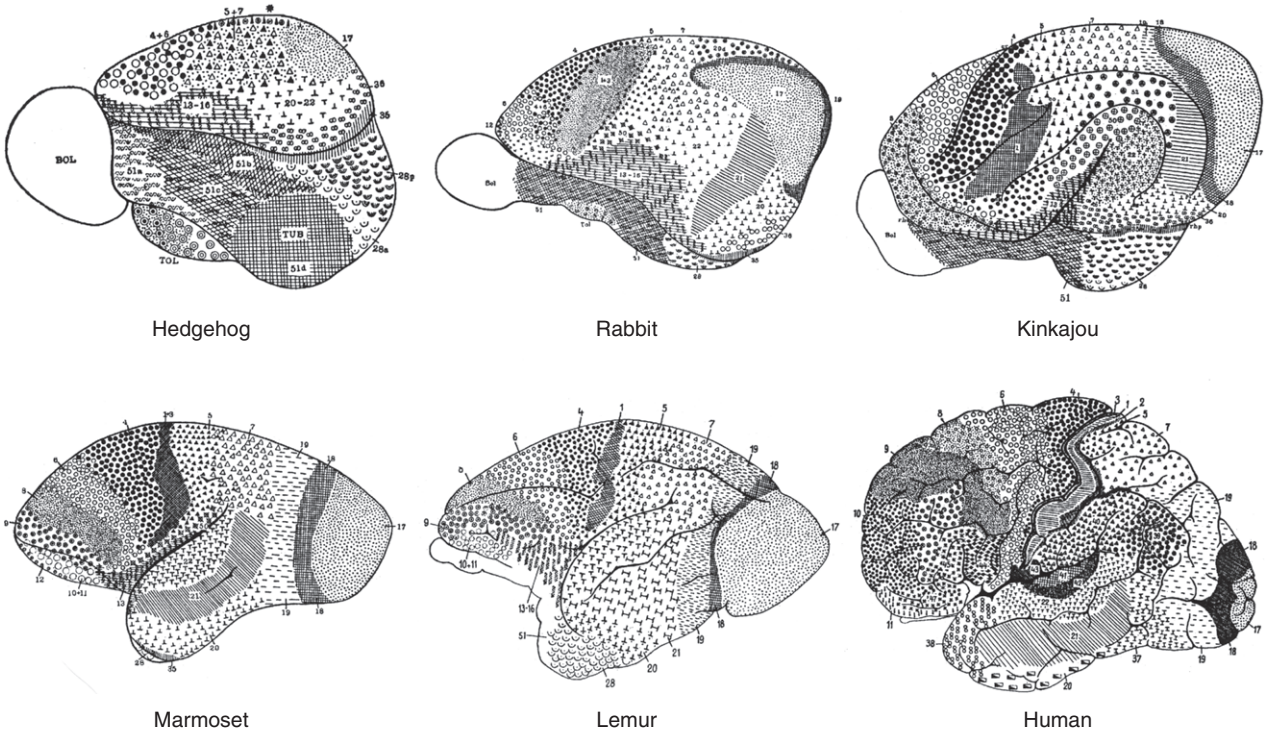


FIGURE 2.21 A similar cerebral cortical regionalization plan for mammals was proposed by Korbinian Brodmann in 1909. His cortical parceling was based on regional differences in how neuronal cell bodies tend to distribute in layers, an approach referred to as cytoarchitectonics. This figure summarizes his findings in six species, with different regions, or "areas," as he called them, indicated with different symbols and numbers. He distinguished 47 areas in the human cerebral cortex and showed that generally similar patterns applied to all nine species he analyzed.

closely to the CNS's surface, even where there are deep invaginations, as in the cerebral and cerebellar cortex. Then comes the arachnoid ("spidery"), which has a tenuous, web-like structure but is histologically similar to pia. Finally, the dura ("tough") is a thick, inelastic covering apposed to the skull and vertebral canal's inner surface (Fig. 2.20*H,O,P*). Membranes covering the CNS are continuous with similar coverings of the PNS, where the terminology differs.

In certain CNS regions neural tissue is absent but meninges persist. Here, ependymal cells lining the ventricular system (the monolayer vestige of the embryonic neuroepithelial layer that ends up lining the adult ventricular system) fuse with the pia and arachnoid layers to form structures known as *choroid plexus*, which contains abundant blood vessels and serves as a component of the blood–brain barrier (the blood–CSF barrier; see Chapter 3). The choroid plexus produces CSF in the roof of the lateral, third, and fourth ventricles, and this fluid fills the brain ventricles, and perhaps at least part of the spinal central canal. Under positive hydrostatic pressure, CSF passes out of the brain's interior through three foramina (holes) in the fourth ventricle's roof, to fill the subarachnoid space between pia and arachnoid.

The Brain is Highly Vascular

The human brain consumes about 20% of the body's oxygen supply at rest, even though it usually weighs only about a kilogram. Thus, the brain must continuously receive a voluminous blood supply, on the order of a liter per minute. Blood reaches the brain through two arterial roots—*vertebral* and *internal carotid arteries*—that anastomose in the circle of Willis, which essentially surrounds the base of the hypothalamus and pituitary gland's stalk. The functional importance of this arterial circle cannot be overemphasized because afterward there is a drastic reduction in anastomoses between major arteries within brain tissue itself. As a result, blockage or rupture of an artery rapidly deprives the supplied brain region of oxygen, causing a *stroke* or brain attack.

After entering the skull through the foramen magnum with the spinal cord, the paired vertebral arteries fuse into a single *basilar artery*, generating the cerebellar arteries and the posterior cerebral arteries, which supply occipital cerebral regions. The internal carotid arteries divide to form the anterior and middle cerebral arteries; the former supplies each hemisphere's medial surface (especially the limbic lobe), and the latter supplies the rest of the hemisphere (including the speech and somatic sensorimotor areas).

By and large, the major arteries course along the cerebral surface and branches dive perpendicularly into the brain and proliferate into arterioles and capillaries.

A system of large *venous sinuses* collect blood from brain capillaries, and CSF from the subarachnoid space, and return them to the heart, mostly via the paired *internal jugular veins*. The major venous sinuses lie within the dura, whose inelasticity essentially holds the sinuses open. Blood flow through sinuses is slow and under low hydrostatic pressure. Thin-walled venous sinuses surrounded by the tough and immovable dura sets the stage for serious injury when the head is subjected to physical trauma. A well-known example is traumatic injury to the great cerebral vein (of Galen) in the midline that can occur when a boxer is struck in the head. The blow's impact causes the brain to recoil in its CSF cushion, exerting a shearing force against the dura, which remains attached to the skull. This force may rupture the great cerebral vein, leading to serious hemorrhage of venous blood into the subdural space between dura and arachnoid.

Summary

This chapter reviews common approaches to the problem of understanding the nervous system's fundamental structure and wiring diagram—the basic plan or architecture. One approach examines a series of increasingly complex animals from an evolutionary perspective to gain insight into basic organizing principles. It reveals trends toward centralization, cephalization, bilateral symmetry, and regionalization of the nervous system. It also suggests that basic molecular and cellular mechanisms of neuronal function, including electrical signal propagation and neurotransmitter release, have changed little since the appearance of the simplest nervous systems in hydra, jellyfish, and other cnidarians.

Another approach follows the vertebrate nervous system's development from embryo to adult. At early developmental stages the CNS of all vertebrates has the same basic structure. A polarized, bilaterally symmetrical, regionalized neural plate of ectodermal origin invaginates to form a neural tube whose rostral half presents three swellings (forebrain, midbrain, and then hindbrain), followed by a caudal presumptive spinal cord. These four basic CNS divisions, arranged from rostral to caudal, go on to subdivide repeatedly until all laminated and nuclear neuron groups (regions) of the adult CNS are formed. A topographic, "geographic," or regional account of the CNS emerges from this developmental approach.

How the CNS's functional systems or circuitry are arranged into a unified whole is a tantalizing, deep, unsolved problem. The model discussed here equates behavior with motor output, which is driven by a combination of sensory, intrinsic behavioral state, and cognitive inputs—as well as by endogenous neuronal activity within each system. Future behavior is determined partly by sensory feedback related to the original behavior's consequences. As this is written, the relationship between CNS macroregionalization (Figs. 2.14 and 2.15) and functional systems (Fig. 2.16) is not obvious. The correspondence between functional neural systems and gene expression networks is even more obscure, although promising results are beginning to emerge in the embryonic spinal cord and brainstem cranial nerve nuclei.

References

Alvarez-Bolado, G., & Swanson, L. W. (1996). *Developmental brain maps: Structure of the embryonic rat brain*. Amsterdam: Elsevier.

Bota, M., & Swanson, L. W. (2010). Collating and curating neuroanatomical nomenclatures: principles and use of the Brain Architecture Knowledge Management System (BAMS). *Front. Neuroinform, 4*(3), 1–16.

Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Barth, Leipzig. Translated as Brodmann's *Localisation in the Cerebral Cortex* by L. J. Garey. London: Gordon-Smith. 1994.

Brusca, R. C., & Brusca, G. J. (1990). *Invertebrates*. Sunderland, MA: Sinauer Associates.

Cajal, S. Ramón y. (1909–1911). *Histologie du système nerveux de l'homme et des vertébrés*, in 2 vols., Maloine, Paris. Translated as *Histology of the Nervous System of Man and Vertebrates* by N. Swanson and L. W. Swanson. New York: Oxford University Press. 1995.

Cartmill, M., Hylander, W. L., & Shafland, J. (1987). *Human structure*. Cambridge, MA: Harvard University Press.

Gazzaniga, M. S. (2005). Forty-five years of split-brain research and still going strong. *Nature Reviews Neuroscience, 6*, 653–659.

Hamburger, V. (1973). Anatomical and physiological basis of embryonic motility in birds and mammals. In G. Gottlieb (Ed.), *Studies on the development of behavior and the nervous system* (Vol. 1, pp. 51076). New York: Academic Press.

Henle, J. (1871). *Handbuch der Nervenlehre des Menschen*. Braunschweig: Vieweg.

Holland, P. W., & Takahashi, T. (2005). The evolution of homeobox genes: Implications for the study of brain development. *Brain Research Bulletin, 66*, 484–490.

Koizumi, O. (2002). Developmental neurobiology of hydra, a model animal of cnidarians. *Canadian Journal of Zoology, 80*, 1678–1689.

Lentz, T. L. (1968). *Primitive nervous systems*. New Haven: Yale University Press.

Luo, L., Callaway, E.M., & Svoboda, K. (2008) Genetic dissection of neural circuits. *Neuron, 57*, 634–660.

McConnell, C. H. (1932). Development of the ectodermal nerve net in the buds of Hydra. *The Quarterly Journal of Microscopical Science, 75*, 495–509.

Nieuwenhuys, R., Voogd, J., & van Huijzen, C. (1988). *The human central nervous system: A synopsis and atlas* (3rd ed.). Berlin: Springer-Verlag.

Parker, G. H. (1919). *The elementary nervous system*. Philadelphia: Lippincott.

Peters, A., Palay, S. L., & Webster, H. deF. (1991). *The fine structure of the nervous system: Neurons and their supporting cells* (3rd ed.). New York: Oxford University Press.

Sherrington, C. S. (1906). *The integrative action of the nervous system*. New York: Scribner's. (Reprinted, Yale University Press, New Haven, 1947).

Sidman, R. L., Angevine, J. B., Jr., & Taber Pierce, E. (1971). *Atlas of the mouse brain and spinal cord*. Cambridge, MA: Harvard University Press.

Singer, C. (1952). *Vesalius on the Human Brain: Introduction*, Translation of the Text, Translation of Descriptions of Figures, Notes to the Translations, Figures. Oxford: Oxford University Press.

Sporns, O., Tonono, G., & Kötter, R. (2005). The human connectome: a structural description of the human brain. *PLoS Computational Biology, 1*(4), 0245–0251.

Swanson, L. W. (1992). *Brain maps: structure of the rat brain*. Amsterdam: Elsevier.

Swanson, L. W. (2000a). What is the brain? *Trends in Neuroscience, 23,* 519–527.

Swanson, L. W. (2000b). A history of neuroanatomical mapping. In A. W. Toga, & J. C. Mazziotta (Eds.), *Brain mapping: The applications* (pp. 77–109). San Diego: Academic Press.

Swanson, L. W. (2012). *Brain architecture: understanding the basic plan* (2nd ed.). Oxford: Oxford University Press.

Von Neumann, J. (1958). *The computer and the brain*. New Haven: Yale University Press.

Wiener, N. (1948). *Cybernetics, or control and communication in the animal and machine*. New York: Wiley.

Williams, P. L. (Ed.), (1995). *Gray's anatomy* (38th (British) Ed.). New York: Churchill Livingstone.

Suggested Readings

Bergquist, H., & Källén, B. (1954). Notes on the early histogenesis and morphogenesis of the central nervous system in vertebrates. *The Journal of Comparative Neurology, 100,* 627–659.

Björklund, A., & Hökfelt, T. (1983–present). *Handbook of chemical neuroanatomy*. Amsterdam: Elsevier.

Bohland, J. W., Wu, C., Barbas, H., Bokil, H., Bota, M., Breiter, H. C., … Mitra, P. P. (2009). A proposal for a coordinated effort for the determination of a brainwide neuroanatomical connectivity in model organisms at a mesoscopic scale. *PLoS Computational Biology, 5*(3), 1–9.

Descartes, R. (1972). *Treatise on Man*. French text with translation by T. S. Steele. Cambridge: Harvard University Press.

Herrick, C. J. (1948). *The brain of the tiger salamander*. Chicago: University of Chicago Press.

Jones, A. R., Overly, C. C., & Sunkin, S. M. (2009). The Allen Brain Atlas: 5 years and beyond. *Nature Reviews Neuroscience, 10,* 1–8.

Kingsbury, B. F. (1922). The fundamental plan of the vertebrate brain. *The Journal of Comparative Neurology, 34,* 461–491.

Lorenz, K. (1978). *Behind the mirror*. Orlando, FL: Harcourt Brace Jovanovich.

Russell, E. S. (1916). *Form and function: A contribution to the history of animal morphology*. London: John Murray.

Swanson, L. W. & Bota, M. (2010). Foundational model of nervous system structural connectivity with a schema for wiring diagrams, connectome, and basic plan architecture. *Proceedings of the National Academy of Sciences of the United States of America, 107,* 20610–20617.

Tinbergen, N. (1951). *The study of instinct*. London: Oxford University Press.

Larry W. Swanson

SECTION II

CELLULAR AND MOLECULAR NEUROSCIENCE

This page intentionally left blank

CHAPTER

3

Cellular Components of Nervous Tissue

Several types of cellular elements are integrated to constitute normally functioning brain tissue. The neuron is the communicating cell, and many neuronal subtypes are connected to one another via complex circuitries, usually involving multiple synaptic connections. Neuronal physiology is supported and maintained by neuroglial cells, which have highly diverse and incompletely understood functions. These include myelination, secretion of trophic factors, maintenance of the extracellular milieu, and scavenging of molecular and cellular debris from it. Neuroglial cells also participate in the formation and maintenance of the blood–brain barrier, a multicomponent structure that is interposed between the circulatory system and the brain substance and that serves as the molecular gateway to brain tissue.

NEURONS

The neuron is a highly specialized cell type and is the essential cellular element in the CNS. All neurological processes are dependent on complex cell–cell interactions among single neurons as well as groups of related neurons. Neurons can be categorized according to their size, shape, neurochemical characteristics, location, and connectivity, which are important determinants of that particular functional role of the neuron in the brain. More importantly, neurons form circuits, and these circuits constitute the structural basis for brain function. *Macrocircuits* involve a population of neurons projecting from one brain region to another region, and *microcircuits* reflect the local cell–cell interactions within a brain region. Detailed analysis of these macro- and microcircuits is essential for understanding neuronal function and dysfunction in the healthy and the diseased brain.

Broadly speaking, therefore, there are five general categories of neurons: inhibitory neurons that make local contacts (e.g., GABAergic interneurons in the cerebral and cerebellar cortex), inhibitory neurons that make distant contacts (e.g., medium spiny neurons of the basal ganglia or Purkinje cells of the cerebellar cortex), excitatory neurons that make local contacts (e.g., spiny stellate cells of the cerebral cortex), excitatory neurons that make distant contacts (e.g., pyramidal neurons in the cerebral cortex), and neuromodulatory neurons that influence neurotransmission, often at large distances. Within these general classes, the structural variation of neurons is systematic, and careful analyses of the anatomic features of neurons have led to various categorizations and to the development of the concept of cell type. The grouping of neurons into descriptive cell types (such as chandelier, double bouquet, or bipolar cells) allows the analysis of populations of neurons and the linking of specified cellular characteristics with certain functional roles.

General Features of Neuronal Morphology

Neurons are highly polarized cells, meaning that they develop distinct subcellular domains that subserve different functions. Morphologically, in a typical neuron, three major regions can be defined: (1) the cell body (*soma* or *perikaryon*), which contains the nucleus and the major cytoplasmic organelles; (2) a variable number of dendrites, which emanate from the perikaryon and ramify over a certain volume of gray matter and that differ in size and shape, depending on the neuronal type; and (3) a single axon, which extends, in most cases, much farther from the cell body than the dendritic arbor (Fig. 3.1). Dendrites may be spiny (as in pyramidal cells) or nonspiny (as in most interneurons), whereas the axon is generally smooth and emits a variable number of branches (collaterals). In vertebrates, many axons are surrounded by an insulating myelin sheath, which

© 2013 Elsevier Inc. All rights reserved.

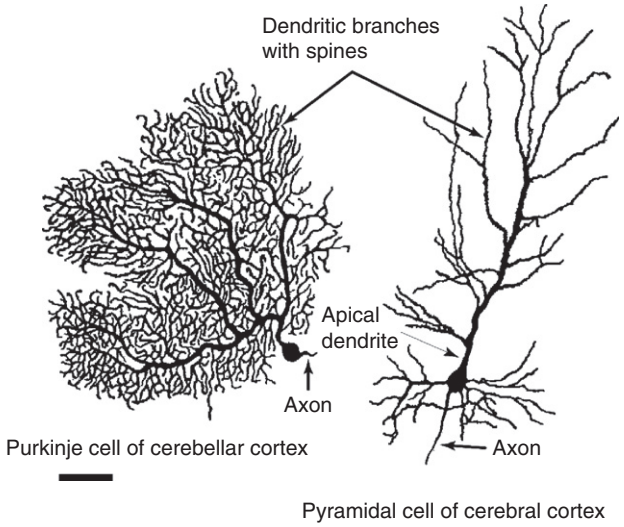


FIGURE 3.1 Typical morphology of projection neurons. (Left) A Purkinje cell of the cerebellar cortex and (right) a pyramidal neuron of the neocortex. These neurons are highly polarized. Each has an extensively branched, spiny apical dendrite, shorter basal dendrites, and a single axon emerging from the basal pole of the cell. The scale bar represents approximately 200 μm.

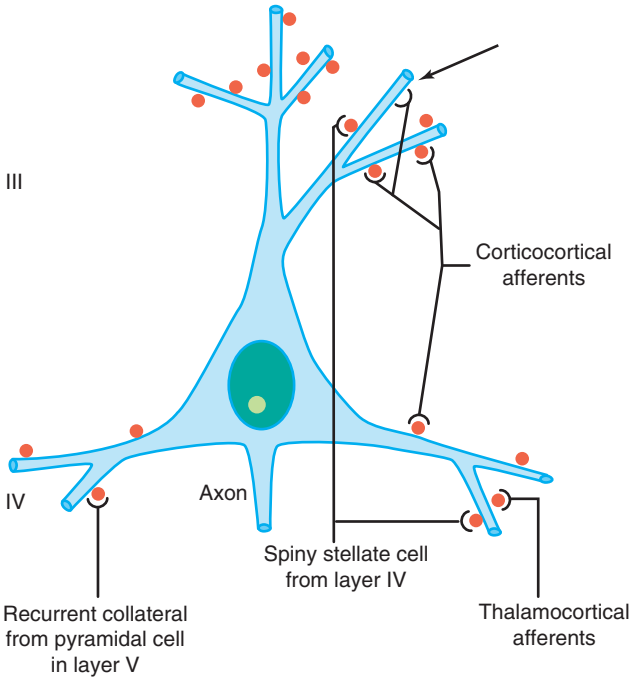


FIGURE 3.2 Schematic representation of four major excitatory inputs to pyramidal neurons. A pyramidal neuron in layer III is shown as an example. Note the preferential distribution of synaptic contacts on spines. Spines are labeled in red. Arrow shows a contact directly on the dendritic shaft.

facilitates rapid impulse conduction. The axon terminal region, where contacts with other cells are made, displays a wide range of morphological specializations, depending on its target area in the central or peripheral nervous system.

The cell body and dendrites are the two major domains of the cell that receive inputs, and dendrites play a critically important role in providing a massive receptive area on the neuronal surface. In addition, there is a characteristic shape for each dendritic arbor, which can be used to classify neurons into morphological types. Both the structure of the dendritic arbor and the distribution of axonal terminal ramifications confer a high level of subcellular specificity in the localization of particular synaptic contacts on a given neuron. The three-dimensional distribution of dendritic arborization is also important with respect to the type of information transferred to the neuron. A neuron with a dendritic tree restricted to a particular cortical layer may receive a very limited pool of afferents, whereas the widely expanded dendritic arborizations of a large pyramidal neuron will receive highly diversified inputs within the different cortical layers in which segments of the dendritic tree are present (Fig. 3.2) (Mountcastle, 1978). The structure of the dendritic tree is maintained by surface interactions between adhesion molecules and, intracellularly, by an array of cytoskeletal components (microtubules, neurofilaments, and associated proteins), which also take part in the movement of organelles within the dendritic cytoplasm.

An important specialization of the dendritic arbor of certain neurons is the presence of large numbers of dendritic spines, which are membranous protrusions. They are abundant in large pyramidal neurons and are much sparser on the dendrites of interneurons (see later).

The perikaryon contains the nucleus and a variety of cytoplasmic organelles. Stacks of rough endoplasmic reticulum are conspicuous in large neurons and, when interposed with arrays of free polyribosomes, are referred to as *Nissl substance*. Another feature of the perikaryal cytoplasm is the presence of a rich cytoskeleton composed primarily of neurofilaments and microtubules, discussed in detail in Chapter 4. These cytoskeletal elements are dispersed in bundles that extend from the soma into the axon and dendrites.

Whereas dendrites and the cell body can be characterized as domains of the neuron that receive afferents, the axon, at the other pole of the neuron, is responsible for transmitting neural information. This information may be primary, in the case of a sensory receptor, or processed information that has already been modified through a series of integrative steps. The morphology of the axon and its course through the nervous system are correlated with the type of information processed by the particular neuron and by its connectivity patterns with other neurons. The axon leaves the cell body from a small swelling called the *axon hillock*. This structure is

particularly apparent in large pyramidal neurons; in other cell types, the axon sometimes emerges from one of the main dendrites. At the axon hillock, microtubules are packed into bundles that enter the axon as parallel fascicles. The axon hillock is the part of the neuron where the action potential is generated. The axon is generally unmyelinated in local circuit neurons (such as inhibitory interneurons), but it is myelinated in neurons that furnish connections between different parts of the nervous system. Axons usually have higher numbers of neurofilaments than dendrites, although this distinction can be difficult to make in small elements that contain fewer neurofilaments. In addition, the axon may show extensive, spatially constrained ramifications, as in certain local circuit neurons; it may give out a large number of recurrent collaterals, as in neurons connecting different cortical regions, or it may be relatively straight in the case of projections to subcortical centers, as in cortical motor neurons that send their very long axons to the ventral horn of the spinal cord. At the interface of axon terminals with target cells are the synapses, which represent specialized zones of contact consisting of a presynaptic (axonal) element, a narrow synaptic cleft, and a postsynaptic element on a dendrite or perikaryon.

Synapses and Spines

Synapses

The exchange of information between neurons mainly takes place through two types of highly specialized structures: chemical synapses (the majority) and electrical synapses. In the latter case, although the plasma membranes of adjacent neurons are separated by a gap of about 2 nm, they contain small channels (gap junctions) that connect the cytoplasm of the adjoining neurons, permitting the diffusion of small molecules and the flow of electric current (Bennett & Zukin, 2004). Each chemical synapse (referred to as "synapse" in the remainder of this chapter) is a complex of several components: (1) a *presynaptic element*, (2) a *cleft*, and (3) a *postsynaptic element*. The presynaptic element is a specialized part of the presynaptic neuron's axon, the postsynaptic element is a specialized part of the postsynaptic somatodendritic membrane, and the space between these two closely apposed elements is the cleft. The portion of the axon that participates in the synapse is the *bouton*, and it is identified by the presence of synaptic vesicles and a presynaptic thickening at the active zone (Fig. 3.3). The postsynaptic element is marked by a postsynaptic thickening opposite the presynaptic thickening. When both sides are equally thick, the synapse is referred to as *symmetric*. When the postsynaptic thickening is greater, the synapse is *asymmetric*. Edward George Gray noticed this difference and divided synapses into two types: *Gray's type 1* synapses are asymmetric and have variably shaped, or pleomorphic, vesicles; *Gray's type 2* synapses are symmetric and have clear, round vesicles. The significance of this distinction is that research has shown that in general, Gray's type 1 synapses tend to be excitatory, whereas Gray's type 2 synapses tend to be inhibitory. This correlation greatly enhanced the usefulness of electron microscopy in neuroscience.

In cross section on electron micrographs, a synapse looks like two parallel lines separated by a very narrow space (Fig. 3.3). Viewed from the inside of the axon or dendrite, it looks like a patch of variable shape. Some synapses are a simple patch, or *macule*. Macular synapses can grow fairly large, reaching diameters over 1 mm. The largest synapses have discontinuities or holes within the macule and are called *perforated synapses* (Fig. 3.3). In cross section, a perforated synapse may resemble a simple macular synapse or several closely spaced smaller macules.

The portion of the presynaptic element that is apposed to the postsynaptic element is the *active zone*. This is the region where the synaptic vesicles are concentrated and where at any time a small number of vesicles are docked, and presumably ready for fusion. The active zone is also enriched with voltage gated calcium channels, which are necessary to permit activity-dependent fusion and neurotransmitter release.

The synaptic cleft is truly a space, but its properties are essential. The width of the cleft (~20 nm) is critical because it defines the volume in which each vesicle releases its contents and, therefore, the peak concentration of neurotransmitter upon release. On the flanks of the synapse, the cleft is spanned by adhesion molecules, which are believed to stabilize the cleft.

The postsynaptic element may be a portion of a soma or a dendrite or, rarely, part of an axon. In the cerebral cortex, most Gray's type 1 synapses are located on dendritic spines, which are specialized protrusions of the dendrite, and most Gray's type 2 synapses are located on somata or dendritic shafts. A similar segregation is seen in cerebellar cortex. In nonspiny neurons, symmetric and asymmetric synapses are often less well separated. Irrespective of location, a postsynaptic thickening marks the postsynaptic element. In Gray's type 1 synapses, the postsynaptic thickening (or postsynaptic density, PSD) is greatly enhanced. Among the molecules that are associated with the PSD are neurotransmitter receptors (e.g., NMDA receptors) and molecules with less obvious function, such as PSD-95.

Spines

Spines are protrusions on the dendritic shafts of some types of neurons and are the sites of synaptic contacts, usually excitatory. Use of the silver impregnation techniques of Golgi or of the methylene blue introduced by Ehrlich in the late nineteenth century led

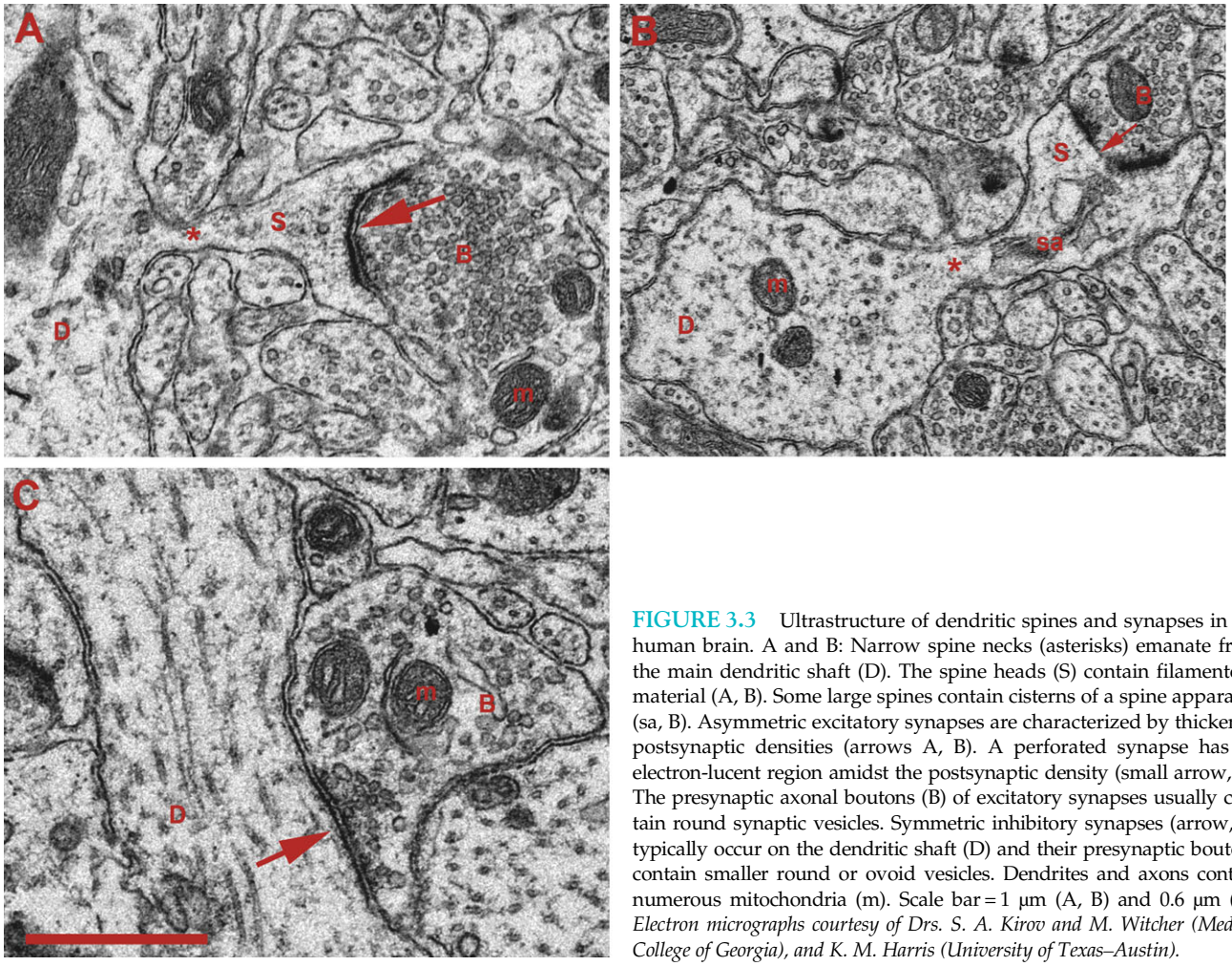


FIGURE 3.3 Ultrastructure of dendritic spines and synapses in the human brain. A and B: Narrow spine necks (asterisks) emanate from the main dendritic shaft (D). The spine heads (S) contain filamentous material (A, B). Some large spines contain cisterns of a spine apparatus (sa, B). Asymmetric excitatory synapses are characterized by thickened postsynaptic densities (arrows A, B). A perforated synapse has an electron-lucent region amidst the postsynaptic density (small arrow, B). The presynaptic axonal boutons (B) of excitatory synapses usually contain round synaptic vesicles. Symmetric inhibitory synapses (arrow, C) typically occur on the dendritic shaft (D) and their presynaptic boutons contain smaller round or ovoid vesicles. Dendrites and axons contain numerous mitochondria (m). Scale bar = 1 μm (A, B) and 0.6 μm (C). *Electron micrographs courtesy of Drs. S. A. Kirov and M. Witcher (Medical College of Georgia), and K. M. Harris (University of Texas–Austin).*

Santiago Ramón y Cajal to the discovery of spiny appendages on dendrites of a variety of neurons. The best known are those on pyramidal neurons and Purkinje cells, although spines occur on neuron types at all levels of the central nervous system. As early as 1890, Santiago Ramón y Cajal suggested that spines could collect the electrical charge resulting from neuronal activity (Cajal, 1909–1911). He also proposed that spines substantially increase the receptive surface of the dendritic arbor, which may represent an important factor in receiving the contacts made by the axonal terminals of other neurons. It has been calculated that the approximately 20,000 spines of a pyramidal neuron account for more than 40% of its total surface area (Peters, Palay, & Webster, 1991).

More recent analyses of spine electrical properties have demonstrated that spines are dynamic structures that can regulate many neurochemical events related to synaptic transmission and modulate synaptic efficacy. Spines are also known to undergo pathologic alterations and have a reduced density in a number of experimental manipulations (such as deprivation of a sensory input) and in many developmental, neurologic, and psychiatric conditions (such as dementing illnesses, chronic alcoholism, schizophrenia, and trisomy 21). Morphologically, spines are characterized by a narrower portion emanating from the dendritic shaft, the neck, and an ovoid bulb or head, although spine morphology may vary from large mushroom-shaped bulbs to small bulges barely discernable on the surface of the dendrite. Spines have an average length of ~2 μm, but there is considerable variability in their dimensions. At the ultrastructural level (Fig. 3.3), spines are characterized by the presence of asymmetric synapses and contain fine and quite indistinct filaments. These filaments most likely consist of actin and α- and β-tubulins. Microtubules and neurofilaments present in dendritic shafts do not enter spines. Mitochondria and free ribosomes are infrequent, although many spines contain polyribosomes in their neck. Interestingly, most polyribosomes in dendrites are located at the bases of spines, where they are associated with endoplasmic reticulum, indicating that spines possess the machinery necessary for the local synthesis of proteins. Another

feature of the spine is the presence of confluent tubular cisterns in the spine head that represent an extension of the dendritic smooth endoplasmic reticulum. Those cisterns are referred to as the *spine apparatus*. The function of the spine apparatus is not fully understood but may be related to the storage of calcium ions during synaptic transmission.

SPECIFIC EXAMPLES OF DIFFERENT NEURONAL TYPES

Inhibitory Local Circuit Neurons

Inhibitory Interneurons of the Cerebral Cortex

The study of cortical interneurons dates back to the original work of Camillo Golgi when he applied the new technique he had discovered, the *reazione nera* (black reaction: Golgi, 1873). This method allowed all the components of a neuron to be visualized in histological sections (soma, dendrites, and axon). Golgi proposed that neurons could generally be considered to be of two different morphological and physiological types: motor (type I) and sensory (type II) neurons. These motor neurons had long axons that sprout collaterals, and they could project beyond the gray matter, whereas by contrast, sensory neurons had short axons that arborized near the parent cell and did not leave the gray matter. This hypothesis was challenged when Cajal entered the scene following his masterful studies of brain structure based on the Golgi method. Cajal argued that it was not physiologically possible to maintain such a distinction, and he designated Golgi's two types as cells with a long axon (projection neurons) and cells with a short axon (intrinsic neurons), avoiding any consideration of their possible physiological roles (Fig. 3.4). Since then, the term short-axon cell has commonly been used as synonymous with *interneurons* (DeFelipe, 2002).

In general, cortical interneurons have been subdivided into two large groups: spiny non-pyramidal cells and aspiny or sparsely spiny non-pyramidal cells. Spiny non-pyramidal cells represent the typical neurons (including spiny stellate cells) of the middle cortical layers (especially layer IV). These spiny non-pyramidal cells are morphologically heterogeneous, with ovoid, fusiform, and triangular somata, and most of them are excitatory (probably glutamatergic). Their axons are found within layer IV or in the layers adjacent to that in which their soma is located, either above or below. Aspiny or sparsely spiny non-pyramidal cells have axons that remain close to the parental cell body, although prominent collaterals may run out from some of these in the horizontal (parallel to the cortical surface) or vertical dimension (ascending and/or descending to other cortical layers). These interneurons constitute approximately 15 to 30% of the total population of cortical neurons, and they appear to be mostly GABAergic, representing the main components of inhibitory cortical circuits (Jones, 1993) (Fig. 3.4). Indeed, they can be found in all cortical layers, and they characteristically display a tremendous variety of morphological, biochemical, and physiological features (Markram et al., 2004; Kubota et al., 2011). However, a general consensus

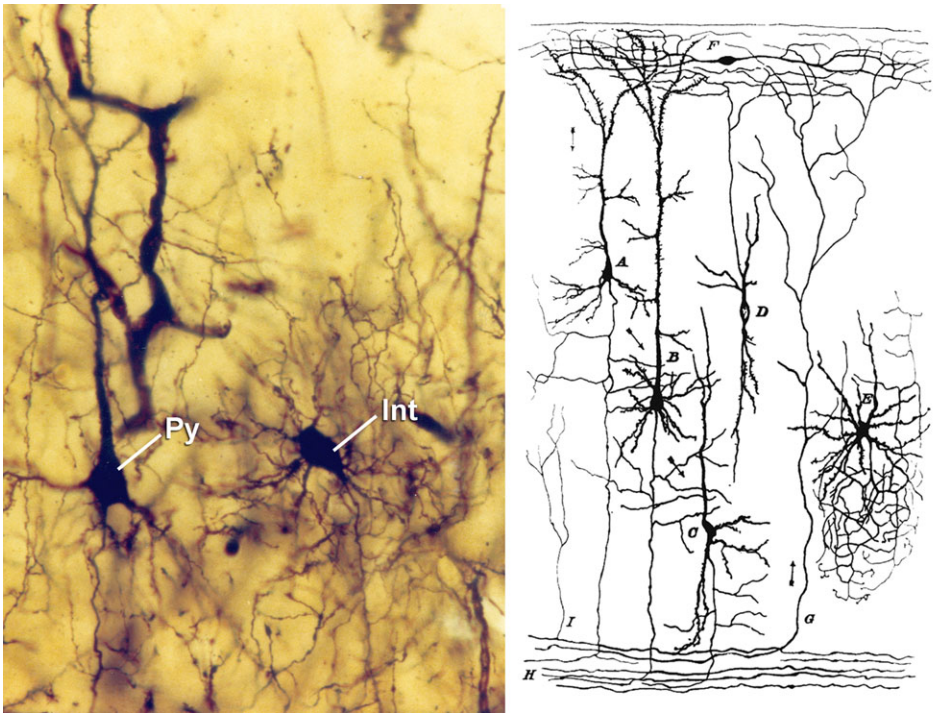


FIGURE 3.4 *Left*, photomicrograph from one of the Cajal's preparations of the "occipital pole" of an eighteen-day-old cat, showing the soma of a pyramidal cell (Py) and a neurogliaform cell (interneuron; Int) stained with the Golgi method. *Right*, the principal cellular types based on the works of Cajal on the cerebral cortex of small mammals (rabbit, guinea pig, rat, and mouse). *A*, pyramidal cell of medium size; *B*, giant pyramidal cell; *C*, polymorphic cell; *D*, cell whose axon is ascending; *E*, cell of Golgi; *F*, special cell of the molecular layer. *Copyright © Herederos de Santiago Ramón y Cajal.*

to name and classify cortical neurons has yet to emerge. This is in part due to the fact that, with some exceptions, there are no rules to establish the essential characteristics that define whether an individual neuron belongs to a given cell type. Indeed, this characterization is also hindered by the fact that most studies of these neurons are generally based on either morphological, physiological, or molecular approaches, rather than any combination of these features.

In a more recent attempt to classify these neurons, the Petilla Interneuron Nomenclature Group (Ascoli et al., 2008) proposed a set of terms to organize and describe the anatomical, molecular, and physiological features of GABAergic interneurons of the cerebral cortex based on existing data. The anatomical, molecular, and physiological classifications divided GABAergic cortical interneurons into several main types of interneurons, which in turn were further subdivided depending on distinct properties. Briefly, the anatomical classification identified three major neuronal types: cells that target pyramidal cells; cells that do not show target specificity; and cells that specifically target other interneurons. The molecular classification includes five main groups based on the expression of specific biochemical markers: parvalbumin, somatostatin, neuropeptide Y in the absence of somatostatin, vasointestinal peptide, and cholecystokinin in the absence of somatostatin and vasointestinal peptide. Finally, the physiological classification identified six main types of interneurons: fast spiking neurons, nonadapting/nonfast spiking neurons, adapting neurons, accelerating cells, irregular spiking neurons, and intrinsic bursting neurons. Each of these attempts to classify these neurons has its limitations, and the scientific community still lacks a general catalogue of accepted neuron types and names. Nevertheless, standardizing the nomenclature based on the properties of GABAergic interneurons proposed by the Petilla Interneuron Nomenclature Group represents a step in the right direction towards a more comprehensive classification of cortical GABAergic interneurons.

In general, GABAergic interneurons control the timing and flow of information in the cerebral cortex (Fig. 3.5), although there does seem to be some division of labor between the different types of interneurons (Klausberger & Somogyi, 2008). One such example is the class of basket and chandelier cells that are considered as fast-spiking GABAergic interneurons and that express the calcium binding protein parvalbumin. These cells appear to play important roles in controlling the timing of pyramidal cell firing, shaping the network output and the rhythms generated in different states of consciousness (Klausberger et al., 2003; Howard, Tamas, & Soltesz, 2005; Yizhar, Fenno,

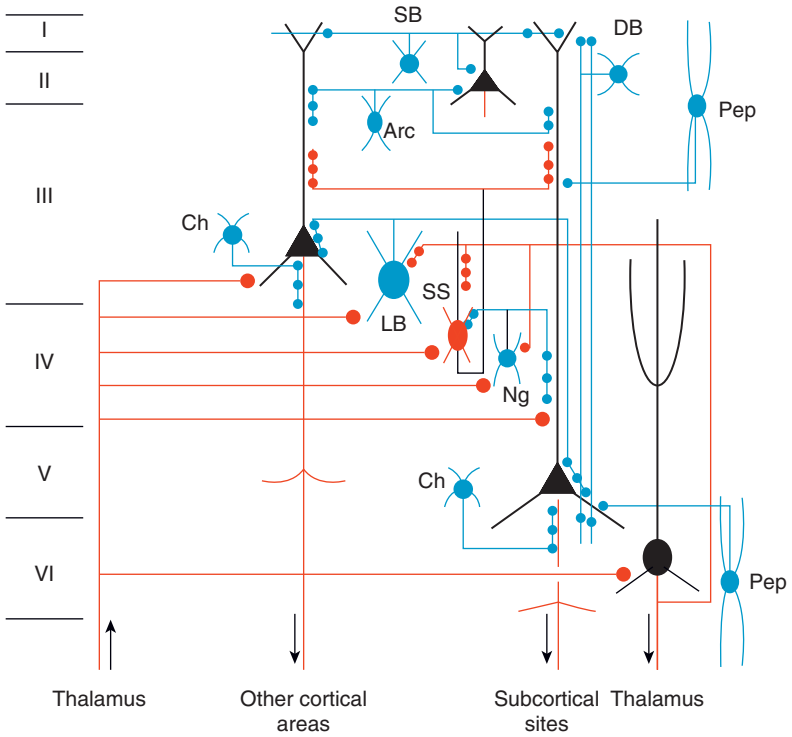


FIGURE 3.5 A schematic circuit based upon the known cortical cells upon which thalamic afferent fibers terminate in cats and monkeys. The GABAergic interneurons (blue) are identified by the names that they have received in these species. *Arc*, neuron with arciform axon; *Ch*, chandelier cell; *DB*, double bouquet cell; *LB*, large basket cell; *Ng*, neurogliaform cell; *Pep*, peptidergic neuron; *SB*, small basket cell. Excitatory neurons (black) include pyramidal cells of layers II–VI and the spiny stellate cells (*SS*) of layer IV. *Courtesy of Dr. E. G. Jones.*

Davidson, Mogri, & Deisseroth, 2011). By contrast, other interneurons containing nitric oxide synthase and various neuropeptides are implicated in the neurovascular coupling (Rossier, 2009). In addition, some types of interneurons have been directly implicated in disease, which is particularly evident when we consider that certain alterations to chandelier and basket cells seem to be critical in establishing some forms of human epilepsy (DeFelipe, 1999; Magloczky & Freund, 2005). Finally, the proportion of GABAergic neurons in the cerebral cortex of rodents is lower than in primates, and there are interneurons in primates that are not found in rodents (Jones, 1975, 1993; DeFelipe, 2002; Markram et al., 2004; Ascoli et al., 2008). These observations, together with the existence of differences in the developmental origins of GABAergic interneurons in rodents and primates, including humans, seem to indicate that more GABA interneurons and newer forms of GABA interneurons appeared in the primate cortex during the course of evolution (Rakic, 2009; Raghanti, Spocter, Butti, Hof, & Sherwood, 2010; DeFelipe, 2011). These features of GABAergic interneurons make their study one of the most exciting and active research fields in relation to the cerebral cortex.

Inhibitory Projection Neurons

Medium-sized Spiny Cells

These neurons are unique to the striatum, a part of the basal ganglia that comprises the caudate nucleus and putamen (see Chapter 31). Medium-sized spiny cells are scattered throughout the caudate nucleus, the putamen, and the nucleus accumbens, and are recognized by their relatively large size, compared with other cellular elements of the basal ganglia. They differ from all other neurons in the striatum in that they have a highly ramified dendritic arborization radiating in all directions, which is densely covered with spines. They furnish a major output from the striatum and receive a highly diverse input from, among other sources, the cerebral cortex, thalamus, and certain dopaminergic neurons of the substantia nigra. These neurons are neurochemically quite heterogeneous, contain GABA, and may contain several neuropeptides and the calcium-binding protein calbindin. In Huntington disease, a neurodegenerative disorder of the striatum characterized by involuntary movements and progressive dementia, an early and dramatic loss of medium-sized spiny cells occurs.

Purkinje Cells

Purkinje cells are the most morphologically striking cellular elements of the cerebellar cortex. They are arranged in a single row throughout the entire cerebellar cortex between the molecular (outer) layer and the granular (inner) layer. Purkinje cells are among the largest neurons in the brain and have a round perikaryon, classically described as shaped "like a chianti bottle," with a highly branched dendritic tree resembling a candelabrum and extending into the molecular layer where they are contacted by incoming systems of afferent fibers from granule neurons and the brainstem (see Chapter 32). Their apical dendrites have an enormous number of spines (more than 80,000 per cell). A particular feature of the dendritic tree of the Purkinje cell is that it is distributed in one plane, perpendicular to the longitudinal axes of the cerebellar folds, and each dendritic arbor determines a separate domain of cerebellar cortex (Fig. 3.1). The axons of Purkinje neurons course through the cerebellar white matter and contact cerebellar nuclei or vestibular nuclei. These neurons contain the inhibitory neurotransmitter GABA and the calcium-binding protein calbindin. Spinocerebellar ataxia, a severe disorder combining ataxic gait and impairment of fine hand movements, accompanied by dysarthria and tremor, has been documented in some families and is related directly to Purkinje cell degeneration.

Excitatory Local Circuit Neurons

Spiny Stellate Cells

Spiny stellate cells are small multipolar neurons with local dendritic and axonal arborizations. These neurons resemble pyramidal cells in that they are the only other cortical neurons with large numbers of dendritic spines, but they differ from pyramidal neurons in that they lack an elaborate apical dendrite. The relatively restricted dendritic arbor of these neurons is presumably a manifestation of the fact that they are high-resolution neurons that gather afferents to a very restricted region of the cortex. Dendrites rarely leave the layer in which the cell body resides. The spiny stellate cell also resembles the pyramidal cell in that it provides asymmetric synapses that are presumed to be excitatory, and is thought to use glutamate as its neurotransmitter (Peters & Jones, 1984).

The axons of spiny stellate neurons have primarily intracortical targets and a radial orientation, and appear to play an important role in forming links among layer IV, the major thalamorecipient layer, and layers III, V, and VI, the major projection layers. The spiny stellate neuron appears to function as a high-fidelity relay of thalamic inputs, maintaining strict topographic organization and setting up initial vertical links of information transfer within a given cortical area (Peters & Jones, 1984).

Excitatory Projection Neurons

Pyramidal Cells

All cortical output is carried by pyramidal neurons, and the intrinsic activity of the neocortex can be viewed

simply as a means of finely tuning and coordinating their output. A pyramidal cell is a highly polarized neuron, with a major orientation axis perpendicular (or orthogonal) to the pial surface of the cerebral cortex. In cross section, the cell body is roughly triangular (Fig. 3.1), although a large variety of morphologic types exist with elongate, horizontal, or vertical fusiform, or inverted perikaryal shapes. Pyramidal cells are the major excitatory type of neurons and use glutamate as their neurotransmitter. A pyramidal neuron typically has a large number of dendrites that emanate from the apex and form the base of the cell body. The span of the dendritic tree depends on the laminar localization of the cell body, but it may, as in giant pyramidal neurons, spread over several millimeters. The cell body and dendritic arborization may be restricted to a few layers or, in some cases, may span the entire cortical thickness (Jones, 1984).

In most cases, the axon of a large pyramidal cell extends from the base of the perikaryon and courses toward the subcortical white matter, giving off several collateral branches that are directed to cortical domains generally located within the vicinity of the cell of origin (as explained later). Typically, a pyramidal cell has a large nucleus, and a cytoplasmic rim that contains, particularly in large pyramidal cells, a collection of granular material chiefly composed of *lipofuscin*. Although all pyramidal cells possess these general features, they can also be subdivided into numerous classes based on their morphology, laminar location, and connectivity with cortical and subcortical regions (Fig. 3.6) (Jones, 1975).

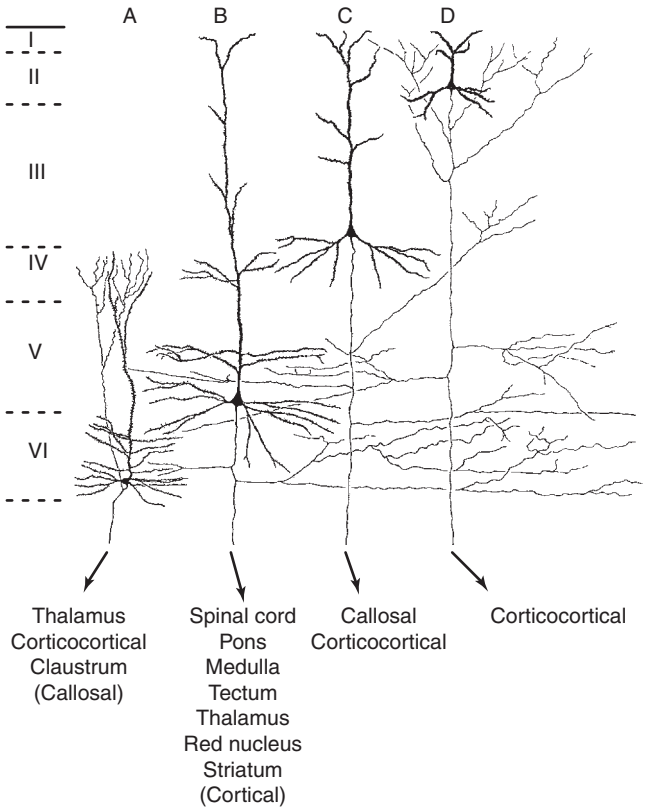


FIGURE 3.6 Morphology and distribution of neocortical pyramidal neurons. Note the variability in cell size and dendritic arborization, as well as the presence of axon collaterals, depending on the laminar localization (I–VI) of the neuron. Also, different types of pyramidal neurons with a precise laminar distribution project to different regions of the brain (less important projections zones are indicated by parentheses). *Adapted from Jones (1984).*

Spinal Motor Neurons

Motor cells of the ventral horns of the spinal cord, also called motoneurons, have their cell bodies within the spinal cord and send their axons outside the central nervous system to innervate the muscles. Different types of motor neurons are distinguished by their targets. The α motoneurons innervate skeletal muscles, but smaller motor neurons (the γ motoneurons, forming about 30% of the motor neurons) innervate the spindle organs of the muscles (see Chapter 28). Alpha motor neurons are some of the largest neurons in the entire central nervous system and are characterized by a multipolar perikaryon and a very rich cytoplasm that renders them very conspicuous in histological preparations. They have a large number of spiny dendrites that arborize locally within the ventral horn. The α motoneuron axon leaves the central nervous system through the ventral root of the peripheral nerves. Their distribution in the ventral horn is not random and corresponds to a somatotopic representation of the muscle groups of the limbs and axial musculature (Brodal, 1981). Spinal motor neurons use acetylcholine as their neurotransmitter. Large motor neurons are severely affected in lower motor neuron disease, a neurodegenerative disorder characterized by progressive muscular weakness that affects, at first, one or two limbs but involves more and more of the body musculature, which shows signs of wasting as a result of denervation.

Neuromodulatory Neurons

Dopaminergic Neurons of the Substantia Nigra

Dopaminergic neurons are large neurons that reside mostly within the pars compacta of the substantia nigra and in the ventral tegmental area (van Domburg & ten Donkelaar, 1991). A distinctive feature of these cells is the presence of a pigment, *neuromelanin*, in compact granules in the cytoplasm. These neurons are medium-sized to large, fusiform, and frequently elongated. They have several large radiating dendrites. The axon emerges from the cell body or from one of the dendrites and projects to large expanses of cerebral cortex and to the basal ganglia. These neurons contain the catecholamine-synthesizing

enzyme *tyrosine hydroxylase*, as well as the monoamine dopamine as their neurotransmitter. Some of them contain both calbindin and calretinin. These neurons are affected severely and selectively in Parkinson disease—a movement disorder different from Huntington disease and characterized by resting tremor and rigidity—and their specific loss is the neuropathologic hallmark of this disorder.

NEUROGLIA

The term *neuroglia*, or "nerve glue," was coined in 1859 by Rudolph Virchow, who conceived of the neuroglia as an inactive "connective tissue" holding neurons together in the central nervous system. The metallic staining techniques developed by Cajal and del Rio-Hortega allowed these two great pioneers to distinguish, in addition to the ependyma lining the ventricles and central canal, three types of supporting cells in the CNS: oligodendrocytes, astrocytes, and microglia. In the peripheral nervous system (PNS), the Schwann cell is the major neuroglial component. These cells are much smaller than the neurons—usually glial cells are 4–10 μm across—yet their size depends on their state of maturation and their possible involvement in a pathological process.

Oligodendrocytes and Schwann Cells Synthesize Myelin

Most brain functions depend on rapid communication between circuits of neurons. As shown in depth later, there is a practical limit to how fast an individual bare axon can conduct an action potential. Organisms developed two solutions for enhancing rapid communication between neurons and their effector organs. In invertebrates, the diameters of axons are enlarged. In vertebrates, the myelin sheath (Fig. 3.7) evolved to permit rapid nerve conduction.

Axon enlargement accelerates action potential propagation in proportion to the square root of axonal diameter. Thus larger axons conduct faster than small ones, but substantial increases in conduction velocity require huge axons. The largest axon in the invertebrate kingdom is the squid giant axon, which is about the thickness of a mechanical pencil lead. This axon conducts the action potential at speeds of 10 to 20 m/s. As the axon mediates an escape reflex, firing must be rapid if the animal is to survive. Bare axons and continuous conduction obviously provide sufficient rates of signal propagation for even very large invertebrates, and many human axons also remain bare. However, in the human brain with 10 billion neurons, axons cannot be as thick as pencil lead, or heads would weigh one hundred pounds or more.

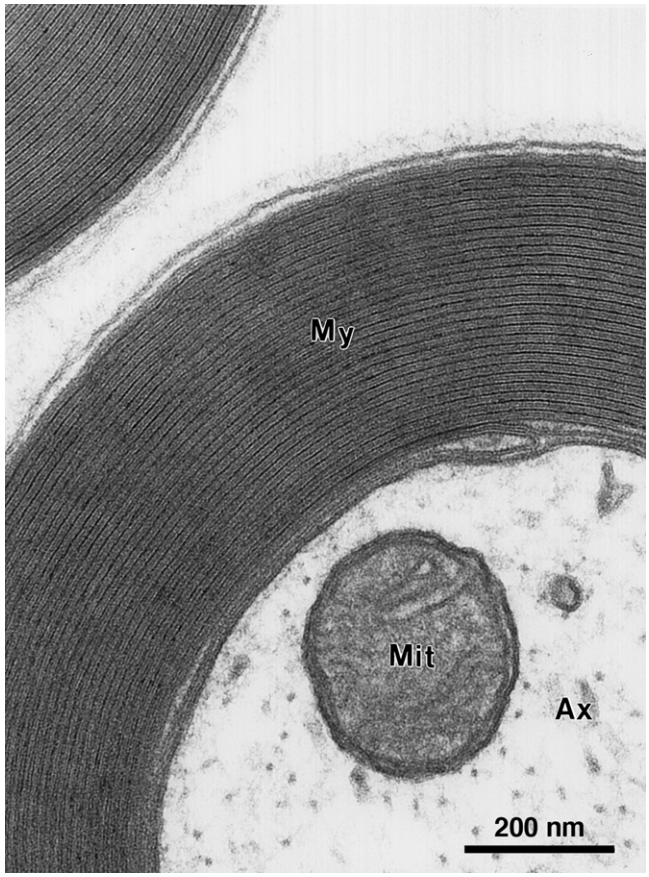


FIGURE 3.7 An electron micrograph of a transverse section through part of a myelinated axon from the sciatic nerve of a rat. The tightly compacted multilayer myelin sheath (My) surrounds and insulates the axon (Ax). Mit, mitochondria. Scale bar = 200 nm.

Thus, along the invertebrate evolutionary line, the use of bare axons imposes a natural, insurmountable limit—a constraint of axonal size—to increasing the processing capacity of the nervous system. Vertebrates, however, get around this problem through evolution of the myelin sheath, which allows 10- to 100-fold increases in conduction of the nerve impulse along axons with fairly minute diameters.

In the central nervous system, myelin sheaths (Fig. 3.8) are elaborated by oligodendrocytes. During brain development, these glial cells send out a few cytoplasmic processes that engage adjacent axons and form myelin around them (Bunge, 1968). Myelin consists of a long sheet of oligodendrocyte plasma membrane, which is spirally wrapped around an axonal segment. At the end of each myelin segment, there is a bare portion of the axon, the node of Ranvier. Myelin segments are thus called *internodes*. Physiologically, myelin has insulating properties such that the action potential can "leap" from node to node and therefore does not have to be regenerated continually along the axonal segment that is covered by the myelin membrane sheath. This leaping of the action

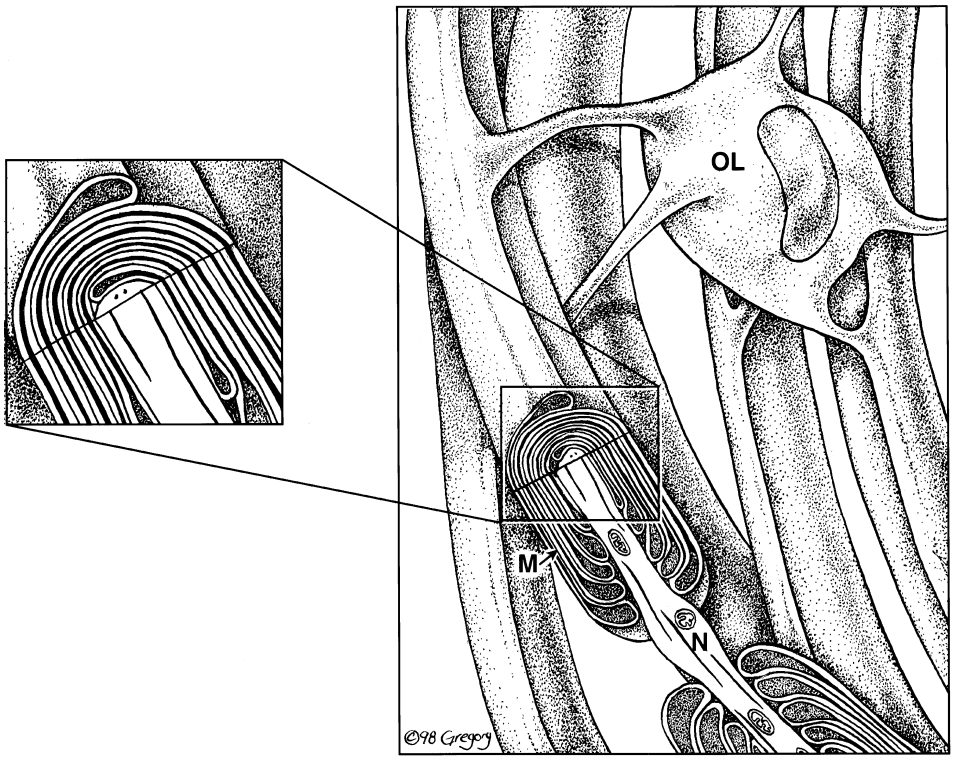


FIGURE 3.8 An oligodendrocyte (OL) in the central nervous system is depicted myelinating several axon segments. A cutaway view of the myelin sheath is shown (M). Note that the internode of myelin terminates in paranodal loops that flank the node of Ranvier (N). (Inset) An enlargement of compact myelin with alternating dark and light electron-dense lines that represent intracellular (major dense lines) and extracellular (intraperiod line) plasma membrane appositions, respectively.

potential from node to node allows axons with fairly small diameters to conduct extremely rapidly (Ritchie, 1984), and is called *saltatory conduction*.

Because the brain and spinal cord are encased in the bony skull and vertebrae, evolution has promoted compactness among the supporting cells of the CNS. Each oligodendrocyte cell body is responsible for the construction and maintenance of several myelin sheaths (Fig. 3.8), thus reducing the number of glial cells required. In both PNS and CNS myelin, cytoplasm is removed between each turn of the myelin, leaving only the thinnest layer of plasma membrane. Due to protein composition differences, CNS lamellae are approximately 30% thinner than in PNS myelin. In addition, there is little or no extracellular space or extracellular matrix between the myelinated axons passing through CNS white matter. Brain volume is thus reserved for further expansion of neuronal populations.

Peripheral nerves pass between moving muscles and around major joints and are routinely exposed to physical trauma. A hard tackle, slipping on an icy sidewalk, or even just occupying the same uncomfortable seating posture for too long can painfully compress peripheral nerves and potentially damage them. Thus, evolutionary pressures shaping the PNS favor robustness and regeneration rather than conservation of space. Myelin in the PNS is generated by Schwann cells (Fig. 3.9), which are different to oligodendrocytes in several ways. Individual myelinating Schwann cells form a single internode.

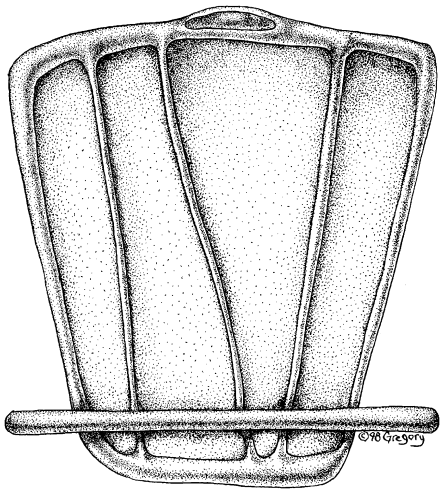


FIGURE 3.9 An "unrolled" Schwann cell in the PNS is illustrated in relation to the single axon segment that it myelinates. The broad stippled region is compact myelin surrounded by cytoplasmic channels that remain open even after compact myelin has formed, allowing an exchange of materials among the myelin sheath, the Schwann cell cytoplasm, and perhaps the axon as well.

The biochemical composition of PNS and CNS myelin differs, as discussed later. Unlike oligodendrocytes, Schwann cells secrete copious extracellular matrix components and produce a basal lamina "sleeve" that runs the entire length of myelinated axons. Schwann cell and fibroblast-derived collagens prevent normal wear-and-tear compression

damage. Schwann cells also respond vigorously to injury, in common with astrocytes but unlike oligodendrocytes. Schwann cell growth factor secretion, debris removal by Schwann cells after injury, and the axonal guidance function of the basal lamina are responsible for the exceptional regenerative capacity of the PNS compared with the CNS.

The major integral membrane protein of peripheral nerve myelin is protein zero (P0), a member of a very large family of proteins termed the immunoglobulin gene superfamily. This protein makes up about 80% of the protein complement of PNS myelin. Interactions between the extracellular domains of P0 molecules expressed on one layer of the myelin sheath with those of the apposing layer yield a characteristic regular periodicity that can be seen by thin section electron microscopy (Fig. 3.7). This zone, called the intraperiod line, represents the extracellular apposition of the myelin bilayer as it wraps around itself. On the other side of the bilayer, the cytoplasmic side, the highly charged P0 cytoplasmic domain probably functions to neutralize the negative charges on the polar head groups of the phospholipids that make up the plasma membrane itself, allowing the membranes of the myelin sheath to come into close apposition with one another. In electron microscopy, this cytoplasmic apposition appears darker than the intraperiod line and is termed the major dense line. In peripheral nerves, although other molecules are present in small quantities in compact myelin and may have important functions, compaction (i.e., the close apposition of membrane surfaces without intervening cytoplasm) is accomplished solely by P0–P0 interactions at both extracellular and intracellular (cytoplasmic) surfaces.

Curiously, P0 is present in the CNS of lower vertebrates such as sharks and bony fish, but in terrestrial vertebrates (reptiles, birds, and mammals), P0 is limited to the PNS. CNS myelin compaction in these higher organisms is subserved by proteolipid protein (PLP) and its alternate splice form, DM-20. These two proteins are generated from the same gene, both span the plasma membrane four times, and differ only in that PLP has a small, positively charged segment exposed on the cytoplasmic surface. Why did PLP/DM-20 replace P0 in CNS myelin? Manipulation of PLP and P0 in CNS myelin established an axonotrophic function for PLP in CNS myelin. Removal of PLP from rodent CNS myelin altered the periodicity of compact myelin and produced a late onset axonal degeneration (Griffiths et al., 1998). Replacing PLP with P0 in rodent CNS myelin stabilized compact myelin but enhanced the axonal degeneration (Yin et al., 2006). These observations indicate that myelination provides trophic support that is essential for axon survival. Studies of primary demyelinating diseases, such as multiple sclerosis, and genetic dysmyelinating diseases (e.g., Charcot-Marie-Tooth diseases) indicate that axonal degeneration is the major cause of permanent disability (Nave & Trapp, 2008).

Myelin membranes also contain a number of other proteins such as the myelin basic protein, which is a major CNS myelin component, and PMP-22, a protein that is involved in a form of peripheral nerve disease. A large number of naturally occurring gene mutations can affect the proteins specific to the myelin sheath and cause neurological disease. In animals, these mutations have been named according to the phenotype that is produced: the shiverer mouse, the shaking pup, the rumpshaker mouse, the jumpy mouse, the myelin-deficient rat, the quaking mouse, and so forth. Many of these mutations are well characterized and have provided valuable insights into the role of individual proteins in myelin formation and axonal survival.

Astrocytes Play Important Roles in CNS Homeostasis

As the name suggests, astrocytes were first described as star-shaped, process-bearing cells distributed throughout the central nervous system. They constitute from 20 to 50% of the volume of most brain areas. Astrocytes appear stellate when stained using reagents that highlight their intermediate filaments but have complex morphologies when their entire cytoplasm is visualized (Fig. 3.10). The two main forms, protoplasmic and fibrous astrocytes, predominate in gray and white matter, respectively (Fig. 3.11). Embryonically, astrocytes develop from radial glial cells, which transversely compartmentalize the neural tube. Radial glial cells serve as scaffolding for the migration of neurons and play a critical role in defining the cytoarchitecture of the CNS (Fig. 3.12). As the CNS matures, radial glia retract their processes and serve as progenitors of astrocytes. However, some specialized astrocytes of a radial nature are still found in the adult cerebellum and the retina and are known as Bergmann glial cells and Müller cells, respectively.

Astrocytes "fence in" neurons and oligodendrocytes. Astrocytes achieve this isolation of the brain parenchyma by extending long processes projecting to the pia mater and the ependyma to form the glia limitans by covering the surface of capillaries and by making a cuff around the nodes of Ranvier. They also ensheath synapses and dendrites and project processes to cell somas (Fig. 3.13). Astrocytes are connected to each other, and to oligodendrocytes, by gap junctions, forming a syncytium that allows ions and small molecules to diffuse across the brain parenchyma. Astrocytes have in common unique cytological and immunological properties that make them easy to identify, including their star shape, the glial end feet on capillaries, and a unique population of large bundles of intermediate filaments. These filaments are composed of an astroglial-specific protein commonly referred to as glial

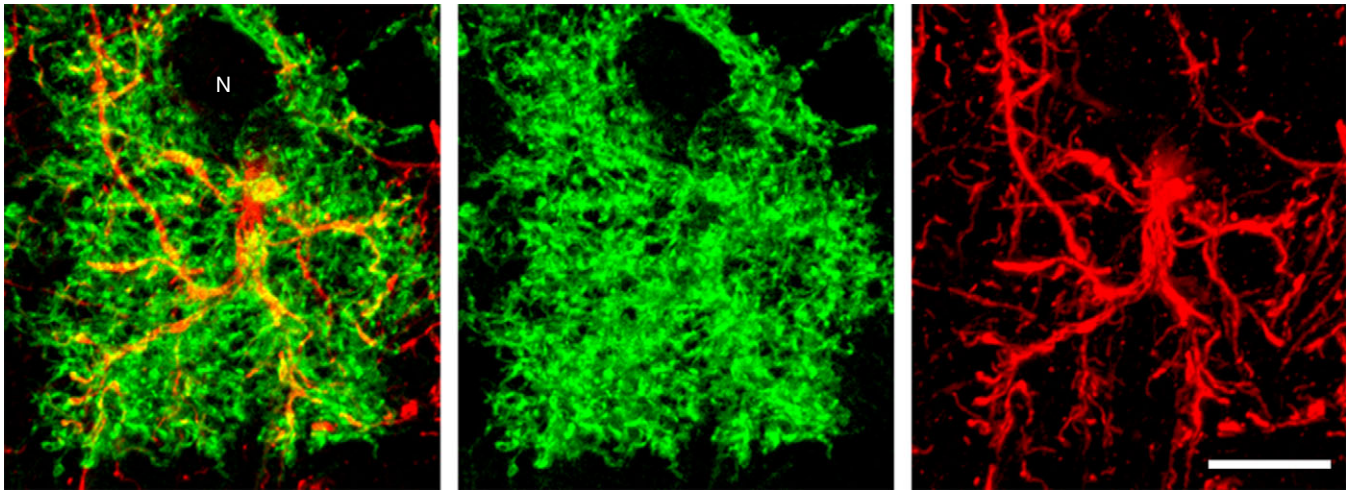


FIGURE 3.10 Astrocytes appear stellate when their intermediate filaments are stained (red, GFAP), but membrane labeling (green, membrane-associated EGFP) highlights the profusion of fine cellular processes that intercalate among other neuropil elements such as synapses and neurons (N). Scale bar = 10 μm. *Image courtesy of Dr. M. C. Smith.*

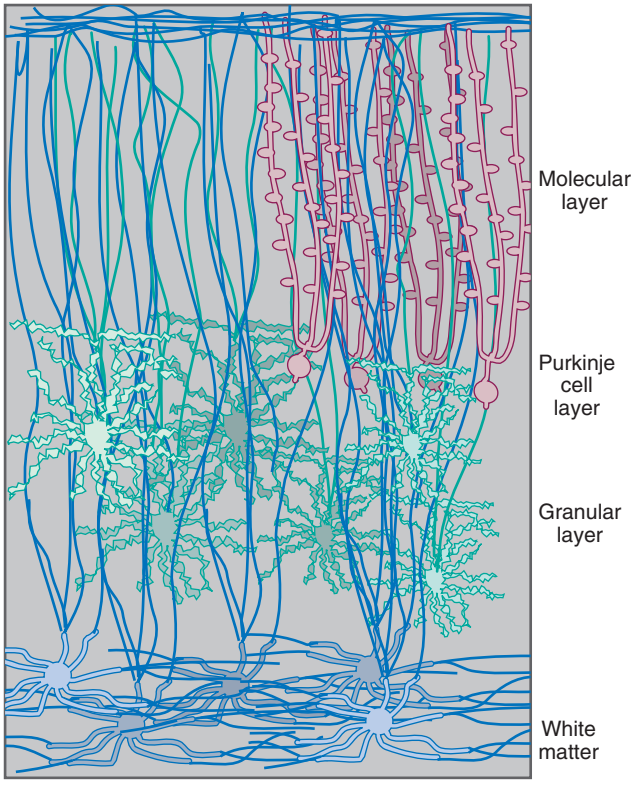


FIGURE 3.11 The arrangement of astrocytes in human cerebellar cortex. Bergmann glial cells are in red, protoplasmic astrocytes are in green, and fibrous astrocytes are in blue.

fibrillary acidic protein (GFAP). S-100, a calcium-binding protein, and glutamine synthetase are also astrocyte markers. Ultrastructurally, gap junctions (connexins), desmosomes, glycogen granules, and membrane orthogonal arrays are distinct features used by morphologists to identify astrocytic cellular processes in the complex cytoarchitecture of the nervous system.

For a long time, astrocytes were thought to physically form the blood–brain barrier (considered later in this chapter), which prevents the entry of cells and diffusion of molecules into the CNS. In fact, astrocytes are indeed the blood–brain barrier in lower species. However, in higher species, astrocytes are responsible for inducing and maintaining the tight junctions in endothelial cells that effectively form the barrier. Astrocytes also take part in angiogenesis, which may be important in the development and repair of the CNS. Their role in this important process is still poorly understood.

Astrocytes Have a Wide Range of Functions

There is strong evidence for the role of radial glia and astrocytes in the migration and guidance of neurons in early development. Astrocytes are a major source of extracellular matrix proteins and adhesion molecules in the CNS; examples are nerve cell–nerve cell adhesion molecule (N-CAM), laminin, fibronectin, cytotactin, and the J-1 family members janusin and tenascin. These molecules participate not only in the migration of neurons, but also in the formation of neuronal aggregates, so-called nuclei, as well as networks.

Astrocytes produce, *in vivo* and *in vitro*, a very large number of growth factors. These factors act singly or in combination to selectively regulate the morphology, proliferation, differentiation, or survival, or all four, of distinct neuronal subpopulations. Most of the growth factors also act in a specific manner on the development and functions of astrocytes and oligodendrocytes. The production of growth factors and cytokines by astrocytes and their responsiveness to these factors is a major

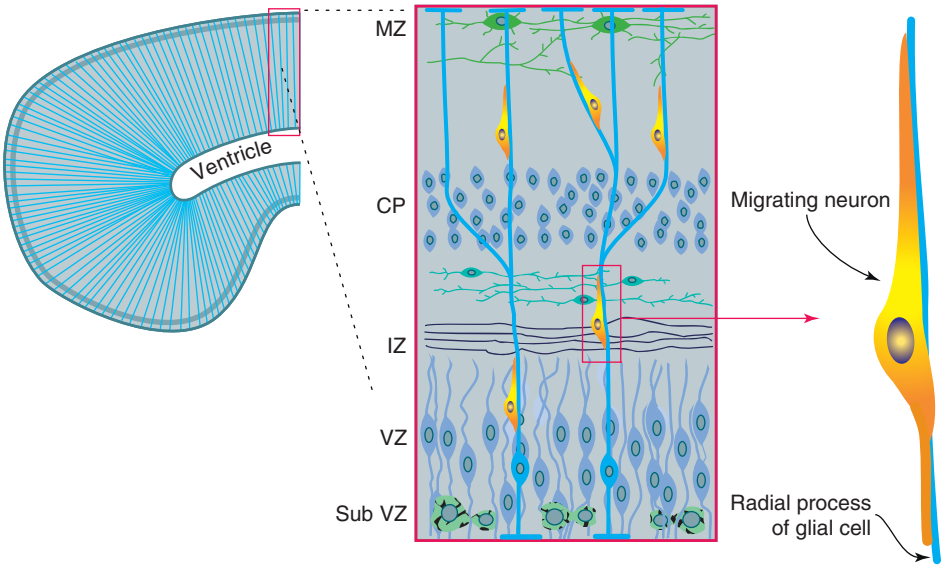


FIGURE 3.12 Radial glia perform support and guidance functions for migrating neurons. In early development, radial glia span the thickness of the expanding brain parenchyma. (Inset) Defined layers of the neural tube from the ventricular to the outer surface: VZ, ventricular zone; IZ, intermediate zone; CP, cortical plate; MZ, marginal zone. The radial process of the glial cell is indicated in blue, and a single attached migrating neuron is depicted at the right.

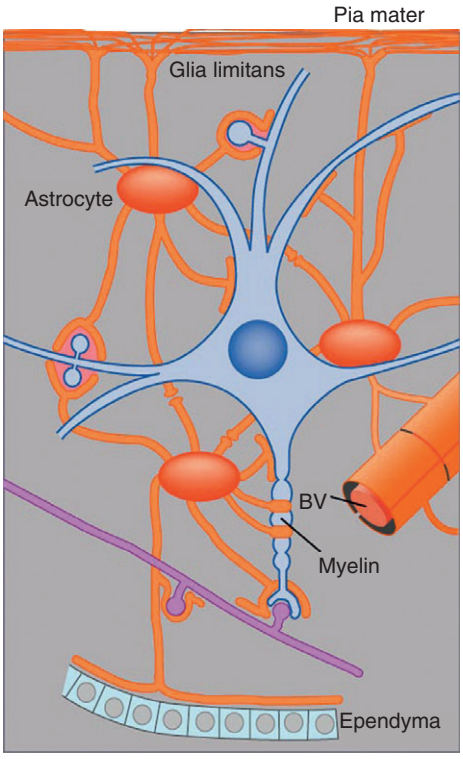


FIGURE 3.13 Astrocytes (in orange) are depicted *in situ* in schematic relationship with other cell types with which they are known to interact. Astrocytes send processes that surround neurons and synapses, blood vessels, and the region of the node of Ranvier and extend to the ependyma, as well as to the pia mater, where they form the glial limitans.

mechanism underlying the developmental function and regenerative capacity of the CNS. During neurotransmission, neurotransmitters and ions are released at high concentration in the synaptic cleft. The rapid removal of these substances is important so that they do not interfere with future synaptic activity. The presence of astrocyte processes around synapses positions them well to regulate neurotransmitter uptake and inactivation (Kettenman & Ransom, 1995). These possibilities are consistent with the presence in astrocytes of transport systems for many neurotransmitters. For instance, glutamate reuptake is performed mostly by astrocytes, which convert glutamate into glutamine and then release it into the extracellular space. Glutamine is taken up by neurons, which use it to generate glutamate and GABA, potent excitatory and inhibitory neurotransmitters, respectively (Fig. 3.14). Astrocytes contain ion channels for K^+, Na^+, Cl^-, HCO_3^-, and Ca^{2+}, as well as displaying a wide range of neurotransmitter receptors. K ions released from neurons during neurotransmission are soaked up by astrocytes and moved away from the area through astrocyte gap junctions. This is known as *spatial buffering*. Astrocytes play a major role in detoxification of the CNS by sequestering metals and a variety of neuroactive substances of endogenous and xenobiotic origin.

In response to stimuli, intracellular calcium waves are generated in astrocytes. Propagation of the Ca^{2+} wave can be visually observed as it moves across the cell soma and from astrocyte to astrocyte. The generation of Ca^{2+} waves from cell to cell is thought to be mediated by second messengers, diffusing through gap junctions (see Chapter 10). In the adult brain, gap junctions are present in all astrocytes. Some gap junctions also have been detected between astrocytes and neurons. Thus, they may participate, along with astroglial neurotransmitter receptors, in the coupling of astrocyte and neuron physiology.

In a variety of CNS disorders—neurotoxicity, viral infections, neurodegenerative disorders, HIV, AIDS, dementia, multiple sclerosis, inflammation, and trauma—astrocytes react by becoming hypertrophic and, in a few

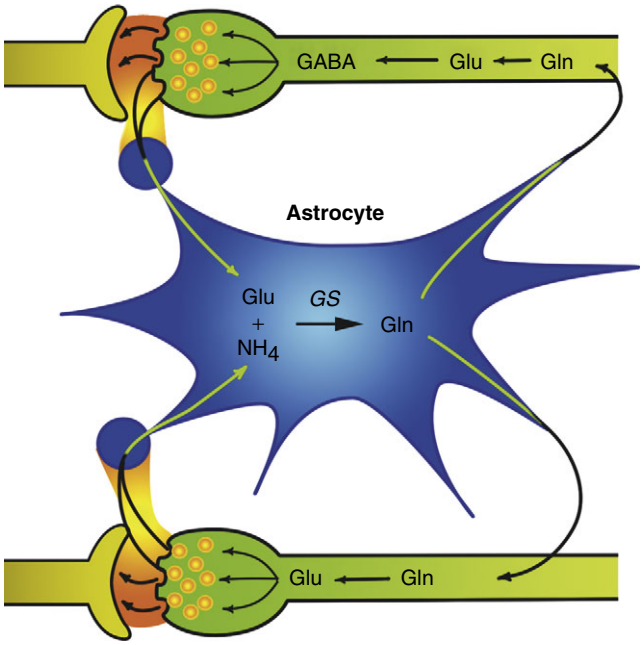


FIGURE 3.14 The glutamate–glutamine cycle is an example of a complex mechanism that involves an active coupling of neurotransmitter metabolism between neurons and astrocytes. The systems of exchange of glutamine, glutamate, GABA, and ammonia between neurons and astrocytes are highly integrated. The postulated detoxification of ammonia and the inactivation of glutamate and GABA by astrocytes are consistent with the exclusive localization of glutamine synthetase in the astroglial compartment.

cases, hyperplastic. A rapid and huge upregulation of GFAP expression and filament formation is associated with astrogliosis. The formation of reactive astrocytes can spread very far from the site of origin. For instance, a localized trauma can recruit astrocytes from as far as the contralateral side, suggesting the existence of soluble factors in the mediation process. Tumor necrosis factor (TNF) and ciliary neurotrophic factors (CNTF) have been identified as key factors in astrogliosis.

Microglia Are Mediators of Immune Responses in Nervous Tissue

The brain traditionally has been considered an "immunologically privileged site," mainly because the blood–brain barrier normally restricts the access of immune cells from the blood. However, it is now known that immunological reactions do take place in the central nervous system, particularly during cerebral inflammation. Microglial cells have been termed the tissue macrophages of the CNS, and they function as the resident representatives of the immune system in the brain. A rapidly expanding literature describes microglia as major players in CNS development and in the pathogenesis of CNS disease.

The first description of microglial cells can be traced to Franz Nissl (1899), who used the term "rod cell" to describe a population of glial cells that reacted to brain pathology. He postulated that rod-cell function was similar to that of leukocytes in other organs. Cajal described microglia as part of his "third element" of the CNS—cells that he considered to be of mesodermal origin and distinct from neurons and astrocytes (Cajal, 1913).

Del Rio-Hortega (1932) distinguished this third element into microglia and oligodendrocytes. He used silver impregnation methods to visualize the ramified appearance of microglia in the adult brain, and he concluded that ramified microglia could transform into cells that were migratory, ameboid, and phagocytic. Indeed, a hallmark of microglial cells is their ability to become reactive and to respond to pathological challenges in a variety of ways. A fundamental question raised by del Rio-Hortega's studies was the origin of microglial cells. Some questions about this remain even today.

Microglia Have Diverse Functions in Developing and Mature Nervous Tissue

On the basis of current knowledge, it appears that most ramified microglial cells are derived from bone marrow–derived monocytes, which enter the brain parenchyma during early stages of brain development. These cells help phagocytosis degenerating cells that undergo programmed cell death as part of normal development. They retain the ability to divide and have the immunophenotypic properties of monocytes and macrophages. In addition to their role in remodeling the CNS during early development, microglia secrete cytokines and growth factors that are important in fiber tract development, gliogenesis, and angiogenesis. They are also the major CNS cells involved in presenting antigens to T lymphocytes. After the early stages of development, ameboid microglia transform into the ramified microglia that persist throughout adulthood (Altman, 1994).

Little is known about microglial function in the healthy adult vertebrate CNS. Microglia constitute a formidable percentage (5–20%) of the total cells in the mouse brain. Microglia are found in all regions of the brain, and there are more in gray than in white matter. The neocortex and hippocampus have more microglia than the brainstem or cerebellum. Species variations also have been noted, as human white matter has three times more microglia than rodent white matter.

Microglia usually have small rod-shaped somas from which numerous processes extend in a rather symmetrical fashion. Processes from different microglia rarely overlap or touch, and specialized contacts between microglia and other cells have not been described in the normal brain. Although each microglial cell occupies its own territory, microglia collectively form a network

that covers much of the CNS parenchyma. Because of the numerous processes, microglia present extensive surface membrane to the CNS environment. Regional variation in the number and shape of microglia in the adult brain suggests that local environmental cues can affect microglial distribution and morphology. On the basis of these morphological observations, it is likely that microglia play a role in tissue homeostasis. The nature of this homeostasis remains to be elucidated. It is clear, however, that microglia can respond quickly and dramatically to alterations in the CNS microenvironment.

Microglia Become Activated in Pathological States

"Reactive" microglia can be distinguished from resting microglia by two criteria: (1) change in morphology and (2) upregulation of monocyte–macrophage molecules (Fig. 3.15). Although the two phenomena generally occur together, reactive responses of microglia can be diverse and restricted to subpopulations of cells within a microenvironment. Microglia not only respond to pathological conditions involving immune activation, but also become activated in neurodegenerative conditions that are not considered immunity-mediated (Graeber, 2010; Naert & Rivest, 2011). This latter response is indicative of the phagocytic role of microglia. Microglia change their morphology and antigen expression in response to almost any form of CNS injury.

CEREBRAL VASCULATURE

The cerebral vasculature supports the development and function of the brain. It is composed of a dense network of arterioles, veinules, and capillaries. Figure 3.16 provides an example of the microvascular density across neocortical layers in a mouse brain. During embryogenesis, the brain vasculature develops via the process of angiogenesis from a preformed perineural vascular plexus. The formation of closed neurovascular networks requires the ordered and coordinated action of neural and endothelial signaling factors, such as vascular endothelial growth factor (VEGF), which is produced in the neuroepithelium and attracts vessel growth into the brain. Other factors, including insulin-like growth factor-1 (IGF-1), basic fibroblast growth factors (bFGFs), interleukin-8 (IL-8), erythropoietin, and angiopoietin-1, promote the recruitment, proliferation, and survival of endothelial cells while simultaneously promoting the proliferation of neural progenitors, neurogenesis, synaptogenesis, and axonal growth. Many neurogenic factors, including nerve growth factor (NGF), brain derived growth factor (BDNF), neuropilin, glial derived growth factor (GDNF), and artemin, as well as components of the extracellular matrix (ECM), affect both neurons and endothelial cells. The vasculature also contributes to neurogenesis by providing a vascular niche for the resident neural progenitor population that exists in neurogenic regions of the adult brain (Ihrie & Alvarez-Buylla, 2011).

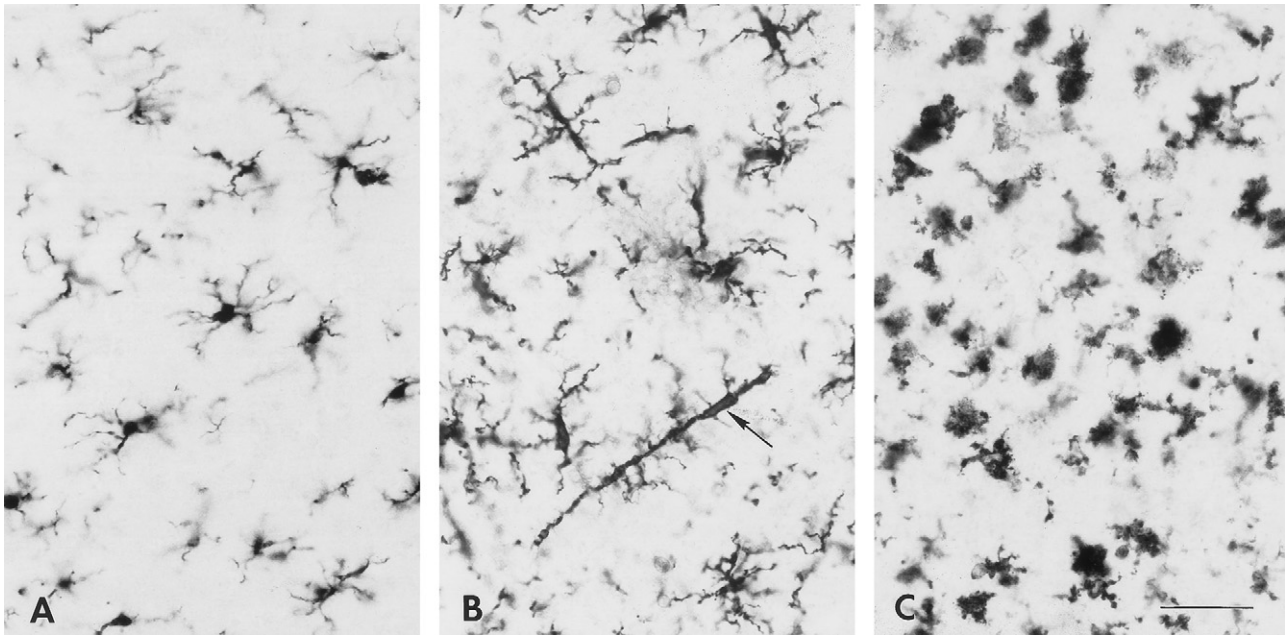


FIGURE 3.15 Activation of microglial cells in a tissue section from human brain. Resting microglia in normal brain (A). Activated microglia in diseased cerebral cortex (B) have thicker processes and larger cell bodies. In regions of frank pathology (C) microglia transform into phagocytic macrophages, which can also develop from circulating monocytes that enter the brain. Arrow in B indicates rod cell. Sections stained with antibody to ferritin. Scale bar = 40 μm.

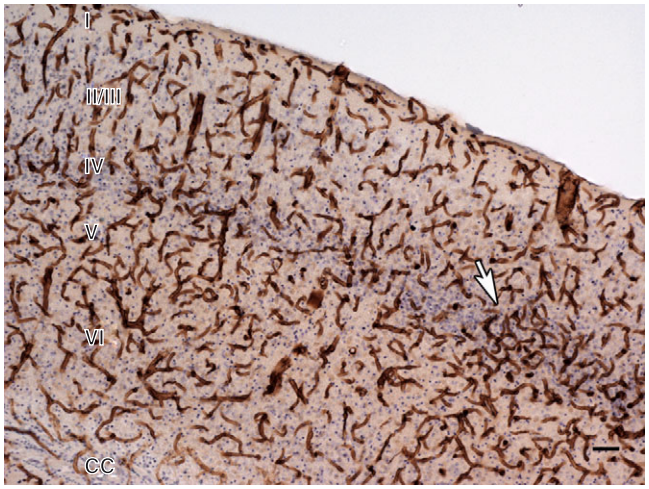


FIGURE 3.16 Microvasculature of the adult mouse somatosensory barrel field (S1BF) cortex. Microvessels were stained with antibodies against collagen type IV, a protein component of the extracellular matrix, and lightly counter-stained with cresyl violet (Nissl). Cortical layers (I–VI) and the corpus callosum (CC) are indicated. An area of increased vascular density in layer IV, where contralateral somatosensory inputs from the thalamus terminate, is indicated by an arrow. Scale bar = 50 μm.

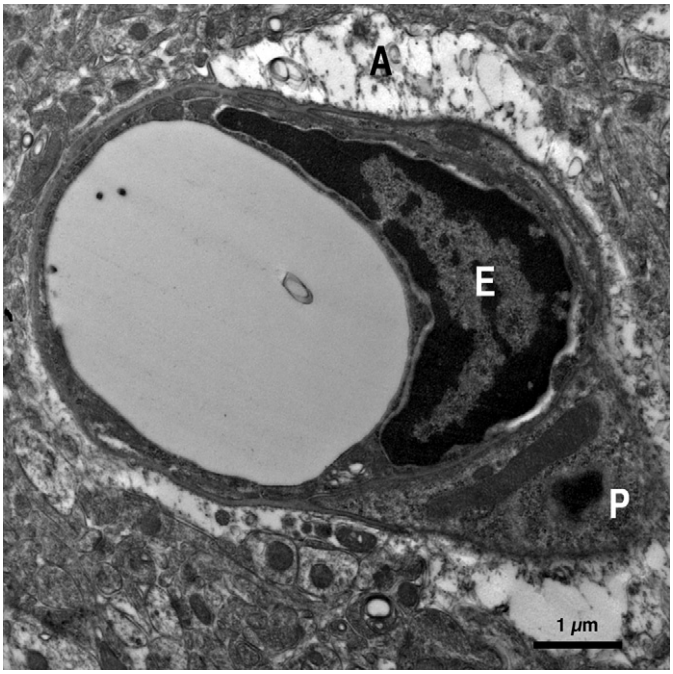


FIGURE 3.17 Ultrastructural analysis of the cerebral microvasculature of a 10-month-old wild-type mouse. A transversely sectioned capillary is shown from the prefrontal cortex. An endothelial cell (E) surrounding the lumen, a pericyte (P), and an astrocytic end-foot process (A) are indicated. Scale bar = 1 μm.

During angiogenesis, endothelial cells proliferate and migrate to form capillary networks, initially aligning into multicellular, precapillary cord-like structures that form an integrated polygonal network. As the vascular cords mature, lumens form and allow blood flow, and endothelial cells become sequestered from the interstitial matrix by a continuous lamina of ECM that forms a basement membrane or basal lamina. The vascular ECM is a hyaline rich and amorphous substance containing laminins, which are thought to be the primary glycoprotein determinants of basement membrane assembly, along with collagen type IV, perlecan (heparan sulfate proteoglycan-2), nidogens, collagen type XVIII, and fibronectin. Structurally, the ECM supports vascular organization as well as playing a mechanosensing role.

Endothelial cells adhere to the ECM through both integrin and nonintegrin receptors. These interactions result in activation of a complex set of signaling pathways that include Rho GTPases (including Cdc42 and Rac1), focal adhesion kinase (FAK), protein kinase Cε (PKCε), and Src. MAPK signaling is particularly important in the regulation of endothelial cell proliferation, migration, and survival. Although these pathways have diverse targets, one common theme is that most affect directly or indirectly the endothelial cytoskeleton. Regulation of ECM structure and function is primarily achieved through the action of matrix metalloproteases (MMPs), which degrade ECM components and release ECM-bound factors such as cytokines that affect vascular cell behavior (Davis, Stratman, Sacharidou, & Koh, 2011).

Transport of Molecules into the Brain

As an extension of the systemic circulation, the cerebral vasculature delivers oxygen and nutrients into the brain and removes carbon dioxide and other metabolic wastes. Endothelial cells interact with neurons, astrocytes, and microglia, as well as other perivascular cells, including smooth muscle cells and pericytes, to form a neurovascular unit (Lecrux & Hamel, 2011) (Fig. 3.17). This structural organization maintains cerebral homeostasis and also forms a blood–brain barrier (BBB) that restricts exchange of solutes between the systemic circulation and brain. Unlike the fenestrated capillaries most often found in the periphery, the spaces between endothelial cells in brain capillaries are occluded by tight junctions (also known as zonula occludens) that form the physical basis for the BBB by restricting the paracellular movement of molecules into the brain. The tight junctions between endothelial cells are localized at cholesterol-rich regions along the plasma membrane containing caveolin-1 (Cav-1) and are composed of tetraspan transmembrane proteins (claudins and occludin) and single-span cytoplasmic proteins (junction adhesion molecules or JAM proteins). Adaptor proteins, such as ZO-1, link both classes of proteins to the actin cytoskeleton as well as to other signaling molecules. These junctions seal the paracellular space, although they remain

capable of rapid modulation and regulation. Consequently, the only solutes that can passively enter the brain are lipid soluble and able to freely diffuse across endothelial cell membranes.

The presence of a functional BBB means that most substances entering the brain must do so by carrier- or receptor-mediated mechanisms. Some molecules cross the BBB by clathrin-mediated endocytosis. However, in addition to clathrin-mediated transport, the plasma membranes of brain endothelia are enriched in membrane lipid rafts, called caveolae, that participate in potocytosis, a type of endocytosis by which small molecules are transported across the plasma membrane by caveolae rather than clathrin-coated vesicles. Select plasma proteins are, for example, taken up and transported across endothelial cell membranes by both receptor-mediated and receptor-independent transcytosis (transport across the interior of the cell from blood to interstitial fluid or vice versa) involving caveolae. Caveolar membranes, in addition to containing structural proteins including the caveolins (Cav-1 and 2 in endothelial cells and Cav-3 in astrocytes) and cavin (abundant at the cytoplasmic face of caveolae), also contain a variety of signaling molecules (G-protein-coupled receptors, G-proteins, non-receptor tyrosine kinases, nonreceptor Ser/Thr kinases, GTPases, and adaptor proteins). Caveolae are also enriched in *β*-D-galactosyl and *β*-*N*-acetylglucosaminyl residues, cholesterol, sphingolipids (sphingomyelin and glycosphingolipids), as well as palmitoleic and stearic acids. Following the BBB breakdown that occurs with acute injuries to the brain, caveolae can fuse to form transendothelial channels that extend from the luminal to the abluminal plasma membrane, which allows passage of macromolecules from the blood to the brain and vice versa, a phenomenon that has not been observed in normal brain endothelium (Nag, Kapadia, & Stewart, 2011; Redzic, 2011).

The high metabolic rate within the brain is dependent on a continuous supply of glucose from the circulation; under normal conditions, glucose is the main source of metabolic energy. Because glucose does not readily cross the BBB, mechanisms have evolved for the transcellular transport of glucose from the blood to the brain interstitial fluid. In the brain, the main glucose transporters are glucose transporter 1 (GLUT-1) in microvascular endothelial cells and glia, and GLUT-3 in neurons. These insulin-insensitive glucose transporters are constitutively expressed on the cell surface. In endothelial cells, the concentration of GLUT-1 is higher on the abluminal surface. Capillary densities in different brain regions are directly related to local blood flow, and close correlations exist between local cerebral glucose utilization, local densities of GLUT-1 and GLUT-3, and local capillary density. Moreover, the pericytes, which are contractile cells, regulate the distribution of the capillary blood flow to match the local cerebral metabolic need (Dalkara, Gursoy-Ozdemir, & Yemisci, 2011).

The BBB also limits the diffusion of other circulating solutes, and independent carrier systems exist for neutral, basic, and acidic amino acids, as well as purines, nucleosides, and monocarboxylic acids, including lactate and pyruvate. Aromatic amino acids serve as precursors for the synthesis of the neurotransmitters serotonin, dopamine, and histamine, and the synthesis of these neurotransmitters is substrate limited. Cerebral endothelial cells express specific solute carrier proteins (SLCs) that mediate the entry and efflux of amino acids across the luminal and abluminal surfaces of the BBB. Both Na^+-dependent and -independent systems have been identified. Na^+-independent amino acid transport systems that have been identified in brain capillary endothelium are system L (or a high-affinity isoform L1) for large neutral amino acids (including L-leucine, L-phenylalanine, L-isoleucine, L-methionine, L-valine); system y^+ for the transport of cationic amino acids (L-arginine, L-lysine, L-ornithine), and system x_c^- for anionic amino acids (cystine/glutamate exchange transporter system). Na^+-dependent transport systems present in brain endothelial cells include systems A and ASC, which show preference for small neutral amino acids (L-alanine, L-serine, L-cysteine), system $B^{o,+}$ for both neutral and basic amino acids, system X_{AG}^- for L-glutamate and L-aspartate, and system *β* (Cl^--dependent) for *β*-amino acids *β*-alanine and taurine. The anionic amino acid transport systems are important in the inactivation of glutamatergic neurotransmission in the brain and for the synthesis of glutathione. The high-affinity Na^+-dependent transporters are located principally in the abluminal membrane of the cerebral endothelium (Broer & Palacin, 2011; Mann, Yudilevich, & Sobrevia, 2003).

The Cerebral Vasculature in Disease States

Many disease states affect the cerebral vasculature. Occlusive cerebral vascular disease can affect both large and small cerebral vessels, causing ischemia of cerebral tissue and resulting acutely in stroke. By contrast, more chronic cerebral ischemia can result in slow, progressive cognitive impairment and vascular dementia. The integrity of the BBB is also affected in a variety of conditions (Nag et al., 2011). For example, brain tumors often disrupt the BBB, causing fluid to leak into brain tissue. The vasogenic edema that results is often associated with neurological impairment beyond the effects of the tumor itself. A variety of inflammatory and infectious conditions also disrupt the BBB. In multiple sclerosis, inflammatory T lymphocytes cross the BBB and initiate an autoimmune attack on central nervous system (CNS) myelin. Treatment modalities often target the effects of an altered BBB, such as the use of corticosteroids in patients with brain tumors or CNS disorders associated with autoimmunity.

References

Altman, J. (1994). Microglia emerge from the fog. *Trends in Neurosciences, 17,* 47–49.

Ascoli, G. A., Alonso-Nanclares, L., Anderson, S. A., Barrionuevo, G., Benavides-Piccione, R., Burkhalter, A., et al. (2008). Petilla Terminology: Nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nature Reviews Neuroscience, 9,* 557–568.

Bennett, M. V., & Zukin, R. S. (2004). Electrical coupling and neuronal synchronization in the Mammalian brain. *Neuron, 41,* 495–511.

Brodal, A. (1981). *Neurological anatomy in relation to clinical medicine* (3rd ed.). New York: Oxford University Press.

Broer, S., & Palacin, M. (2011). The role of amino acid transporters in inherited and acquired diseases. *The Biochemical Journal, 436,* 193–211.

Bunge, R. P. (1968). Glial cells and the central myelin sheath. *Physiological Reviews, 48,* 197–251.

Cajal, S. R. (1909–1911). *Histologie du Système Nerveux de l'Homme et des Vertébrés*. Paris, France: Maloine.

Cajal, S. R. (1913). Contribucion al conocimiento de la neuroglia del cerebro humano. *Trabajos del Laboratorio de Investigaciones Biologicas, 11,* 255–315.

Dalkara, T., Gursoy-Ozdemir, Y., & Yemisci, M. (2011). Brain microvascular pericytes in health and disease. *Acta Neuropathologica, 122,* 1–9.

Davis, G. E., Stratman, A. N., Sacharidou, A., & Koh, W. (2011). Molecular basis for endothelial lumen formation and tubulogenesis during vasculogenesis and angiogenic sprouting. *International Review of Cell and Molecular Biology, 288,* 101–165.

Del Rio-Hortega, P. (1932). Microglia. In W. Penfield (Ed.), *Cytology and cellular pathology of the nervous system* (Vol. 2, pp. 481–534). New York: Harper (Hoeber).

DeFelipe, J. (1999). Chandelier cells and epilepsy. *Brain, 122,* 1807–1822.

DeFelipe, J. (2002). Cortical interneurons: from Cajal to 2001. In E. Azmitia, J. DeFelipe, E. G. Jones, P. Rakic, C. Ribak (Eds.), *Progress in Brain Research: Changing Views of Cajal's Neuron* (Vol. 136, pp. 215–238). Elsevier.

DeFelipe, J. (2011). The evolution of the brain, the human nature of cortical circuits, and intellectual creativity. *Frontiers in Neuroanatomy, 5,* 29. doi: 10.3389/fnana.2011.00029.

Golgi, C. (1873). Sulla struttura della sostanza grigia del cervello (Comunicazione preventiva). *Gazzetta Medica Italiana. Lombardia, 33,* 244–246.

Graeber, M. B. (2010). Changing face of microglia. *Science, 330,* 783–788.

Griffiths, I., Klugmann, M., Anderson, T., Yool, D., Thomson, C., Schwab, M. H., et al. (1998). Axonal swellings and degeneration in mice lacking the major proteolipid of myelin. *Science, 280,* 1610–1613.

Howard, A., Tamas, G., & Soltesz, I. (2005). Lighting the chandelier: new vistas for axo-axonic cells. *Trends in Neurosciences, 28,* 310–316.

Ihrie, R. A., & Alvarez-Buylla, A. (2011). Lake-front property: a unique germinal niche by the lateral ventricles of the adult brain. *Neuron, 70,* 674–686.

Jones, E. G. (1975). Varieties and distribution of non-pyramidal cells in the somatic sensory cortex of the squirrel monkey. *The Journal of Comparative Neurology, 160,* 205–267.

Jones, E. G. (1984). Laminar distribution of cortical efferent cells. In A. Peters & E. G. Jones (Eds.), *Cellular components of the cerebral cortex* (Vol. 1, pp. 521–553). New York: Plenum Press.

Jones, E. G. (1993). GABAergic neurons and their role in cortical plasticity in primates. *Cerebral Cortex, 3,* 361–372.

Kettenman, H., & Ransom, B. R. (Eds.). (1995). *Neuroglia*. Oxford: Oxford University Press.

Klausberger, T., & Somogyi, P. (2008). Neuronal diversity and temporal dynamics: the unity of hippocampal circuit operations. *Science, 321,* 53–57.

Klausberger, T., Magill, P. J., Marton, L. F., Roberts, J. D., Cobden, P. M., Buzsaki, G., & Somogyi, P. (2003). Brain-state- and cell-type-specific firing of hippocampal interneurons in vivo. *Nature, 421,* 844–848.

Kubota, Y., Shigematsu, N., Karube, F., Sekigawa, A., Kato, S., Yamaguchi, N., et al. (2011). Selective coexpression of multiple chemical markers defines discrete populations of neocortical GABAergic neurons. *Cerebral Cortex, 21,* 1803–1817.

Lecrux, C. & Hamel, E. (2011). The neurovascular unit in brain function and disease. *Acta Physiologica*.

Magloczky, Z., & Freund, T. F. (2005). Impaired and repaired inhibitory circuits in the epileptic human hippocampus. *Trends in Neurosciences, 28,* 334–340.

Mann, G. E., Yudilevich, D. L., & Sobrevia, L. (2003). Regulation of amino acid and glucose transporters in endothelial and smooth muscle cells. *Physiological Reviews, 83,* 183–252.

Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., & Wu, C. (2004). Interneurons of the neocortical inhibitory system. *Nature Reviews Neuroscience, 5,* 793–807.

Mountcastle, V. B. (1978). An organizing principle for cerebral function: The unit module and the distributed system. In V. B. Mountcastle & G. Eddman (Eds.), *The mindful brain: cortical organization and the group-selective theory of higher brain function* (pp. 7–50). Cambridge, MA: MIT Press.

Naert, G., & Rivest, S. (2011). The role of microglial cell subsets in Alzheimer's disease. *Current Alzheimer Research, 8,* 151–155.

Nag, S., Kapadia, A., & Stewart, D. J. (2011). Review: molecular pathogenesis of blood-brain barrier breakdown in acute brain injury. *Neuropathology and Applied Neurobiology, 37,* 3–23.

Nave, K. A., & Trapp, B. D. (2008). Axon-glial signaling and the glial support of axon function. *Annual Review of Neuroscience, 31,* 535–561.

Nissl, F. (1899). Über einige Beziehungen zwischen Nervenzellenerkränkungen und gliösen Erscheinungen bei verschiedenen Psychosen. *Archiv für Psychologie, 32,* 1–21.

Peters, A., & Jones, E. G. (Eds.). (1984). *Cellular components of the cerebral cortex* (Vol. 1). New York: Plenum.

Peters, A., Palay, S. L., & Webster, H. deF. (1991). *The fine structure of the nervous system: neurons and their supporting cells* (3rd ed.). New York: Oxford University Press.

Raghanti, M. A., Spocter, M. A., Butti, C., Hof, P. R., & Sherwood, C. C. (2010). A comparative perspective on minicolumns and inhibitory GABAergic interneurons in the neocortex. *Frontiers in Neuroanatomy, 4,* 3. doi: 10.3389/neuro.05.003.2010.

Rakic, P. (2009). Evolution of the neocortex: a perspective from developmental biology. *Nature Reviews Neuroscience, 10,* 724–735.

Redzic, Z. (2011). Molecular biology of the blood–brain and the blood–cerebrospinal fluid barriers: similarities and differences. *Fluids and Barriers of the CNS, 8,* 3.

Ritchie, J. M. (1984). Physiological basis of conduction in myelinated nerve fibers. In P. Morell (Ed.), *Myelin* (pp. 117–146). New York: Plenum.

Rossier, J. (2009). Wiring and plumbing in the brain. *Frontiers in Human Neuroscience, 3,* 2. doi: 10.3389/neuro.09.002.2009.

van Domburg, P. H. M. F., & ten Donkelaar, H. J. (1991). The human substantia nigra and ventral tegmental area. *Advances in Anatomy, Embryology, and Cell Biology, 121,* 1–132.

Yin, X., Baek, R. C., Kirschner, D. A., Peterson, A., Fujii, Y., Nave, K. A., et al. (2006). Evolution of a neuroprotective function of central nervous system myelin. *Journal of Cell Biology, 172,* 469–478.

Yizhar, O., Fenno, L. E., Davidson, T. J., Mogri, M., & Deisseroth, K. (2011). Optogenetics in neural systems. *Neuron, 71,* 9–34.

Suggested Readings

Allaman, I., Bélanger, M., & Magistretti, P. J. (2011). Astrocyte-neuron metabolic relationships: for better and for worse. *Trends in Neurosciences, 34,* 76–87.

Brightman, M. W., & Reese, T. S. (1969). Junctions between intimately apposed cell membranes in the vertebrate brain. *Journal of Cell Biology, 40,* 648–677.

Broadwell, R. D., & Salcman, M. (1981). Expanding the definition of the BBB to protein. *Proceedings of the National Academy of Sciences of the USA, 78,* 7820–7824.

Cardoso, F. L., Brites, D., & Brito, M. A. (2010). Looking at the blood–brain barrier: molecular anatomy and possible investigation approaches. *Brain Research Reviews, 64,* 328–363.

Gehrmann, J., Matsumoto, Y., & Kreutzberg, G. W. (1995). Microglia: Intrinsic immune effector cell of the brain. *Brain Research Reviews, 20,* 269–287.

Kimbelberg, H., & Norenberg, M. D. (1989). Astrocytes. *Scientific American, 26,* 66–76.

Lum, H., & Malik, A. B. (1994). Regulation of vascular endothelial barrier function. *The American Journal of Physiology, 267,* L223–L241.

Nave, K. A. (2010). Myelination and the trophic support of long axons. *Nature Reviews Neuroscience, 11,* 275–283.

Rosenbluth, J. (1999). A brief history of myelinated nerve fibers: one hundred and fifty years of controversy. *Journal of Neurocytology, 28,* 251–262.

Patrick R. Hof, Grahame Kidd, Javier DeFelipe, Jean de Vellis, Miguel A. Gama Sosa, Gregory A. Elder, and Bruce D. Trapp

This page intentionally left blank

CHAPTER

4

Subcellular Organization of the Nervous System: Organelles and Their Functions

Cells have many features in common, but each cell type also possesses a functional architecture related to its unique physiology. In fact, cells may become so specialized in fulfilling a particular function that virtually all subcellular components may be devoted to that function. For example, the machinery inside mature mammalian erythrocytes is completely dedicated to the delivery of oxygen to the tissues and the removal of carbon dioxide. Toward this end, this cell has evolved a specialized plasma membrane, an underlying cytoskeletal matrix that molds the cell into a biconcave disk, and a cytoplasm rich in hemoglobin. Modification of the cell machinery extends even to the discarding of structures such as the nucleus and the protein synthetic apparatus, which are not needed after the red blood cell matures. In many respects, the terminally differentiated, highly specialized cells of the nervous system exhibit comparable commitment; the extensive development of subcellular components reflects the roles that each plays (Brady & Tai, 2012).

The neuron serves as the cellular correlate of information processing, and, in aggregate, all neurons act together to integrate responses of the entire organism to the external world. It is therefore not surprising that the specializations found in neurons are more diverse and complex than those found in any other cell type. Single neurons commonly interact in specific ways with hundreds of other cells—other neurons, astrocytes, oligodendrocytes, immune cells, muscle, and glandular cells (see Chapter 3). Some of these other cell types in the nervous system, such as the myelinating glia (oligodendrocytes and Schwann cells) or astrocytes are inextricably linked to the neuron both functionally and structurally, representing critical partners in building and maintaining a nervous system, with highly specialized subcellular architecture. Here, we define the major functional domains of the neuron, describe the subcellular elements that compose the building blocks of these domains, and examine the processes that create and maintain neuronal functional architecture.

AXONS AND DENDRITES: UNIQUE STRUCTURAL COMPONENTS OF NEURONS

Neurons and glial cells are remarkable for their size and complexity, but they do share many features with other eukaryotic cells (Peters, Palay, & Webster, 1991). As discussed in Chapter 3, the perikaryon, or cell body, contains a nucleus and its associated protein synthetic machinery. Most neuronal nuclei are large and typically contain a preponderance of euchromatin. This is consistent with the need to create and maintain a large cellular volume. Because protein synthesis must be kept at a high level just to maintain the extensive volume and surface area comprising the perikaryon, dendrites, and axon, transcription levels in neurons are generally high. In turn, the wide variety of different polypeptide constituents associated with cellular domains in a neuron requires that a large number of different genes be transcribed constantly.

When specific mRNAs have been synthesized and processed, they move from the nucleus into a subcellular region that can be termed the translational cytoplasm (Lasek & Brady, 1982) comprising cytoplasmic ("free") and membrane-associated polysomes, the intermediate compartment of the smooth endoplasmic reticulum, and

© 2013 Elsevier Inc. All rights reserved.

the Golgi complex. The constituents of translational cytoplasm are thus associated with the synthesis and processing of proteins. Neurons in particular have relatively large amounts of translational cytoplasm to accommodate a high level of protein synthesis. This protein synthetic machinery is arranged in discrete intracellular "granules" termed Nissl substance, after the histologist who first discovered these structures in the nineteenth century. The Nissl substance is actually a combination of stacks of rough endoplasmic reticulum (RER), interposed with rosettes of free polysomes (Fig. 4.1). This arrangement is unique to neurons, and its functional significance is unknown. Most, but by no means all, proteins used throughout the neuron are synthesized in the perikaryon. During or after synthesis and processing, proteins are packaged into membrane-limited organelles in the Golgi complex, incorporated into cytoskeletal elements, or remain as soluble constituents of the cytoplasm. After proteins have been packaged appropriately, they are transported to their sites of function.

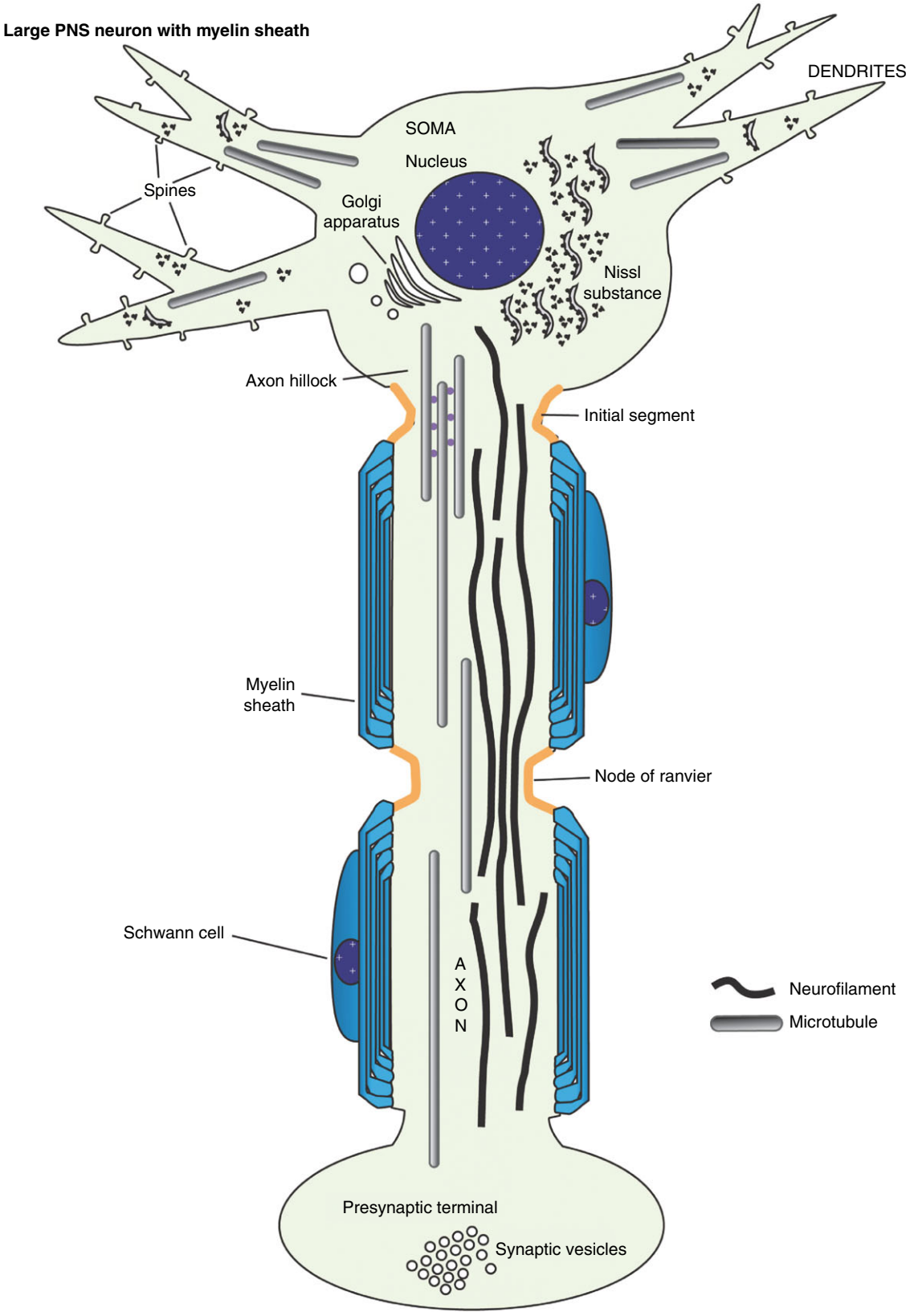
With a few exceptions, vertebrate neurons have two discrete functional domains or compartments, the axonal and the somatodendritic compartments, each of which encompasses a number of distinct microdomains. The axon is perhaps the most familiar functional domain of a neuron and is classically defined as the cellular process by which a neuron makes contact with a target cell to transmit information, providing a conducting structure for transmitting the action potential to a synapse, a specialized subdomain for transmission of a signal from neuron to target cell (neuron, muscle, etc.), most often by release of appropriate neurotransmitters. Consequently, most axons end in presynaptic terminal specializations, although a single axon may have many (hundreds or even thousands in some cases) presynaptic specializations known as en passant synapses along its length. Characteristics of presynaptic terminals are presented in greater detail later.

The axon is the first neuronal process to differentiate during development. A typical neuron has only a single axon that proceeds some distance from the cell body before branching extensively. Usually the longest process of a neuron, axons come in many sizes. In a human adult, axons range in length from a few micrometers for small interneurons to a meter or more for large motor or sensory neurons, and they may be even longer in large animals (such as giraffes, elephants, and whales). In mammals and other vertebrates, the longest axons generally extend approximately half the body length.

Axonal diameters also are quite variable, ranging from 0.1 to 20 μm for large myelinated fibers in vertebrates. Invertebrate axons grow to even larger diameters, with the giant axons of some squid species achieving diameters in the millimeter range. Invertebrate axons reach such large diameters because they lack the myelinating glia that speed conduction of the action potential. As a result, axonal caliber must be large to sustain the high rate of conduction needed for the reflexes that permit escape from predators and capture of prey. Although axonal caliber is closely regulated in both myelinated and nonmyelinated fibers, this parameter is critical for those organisms that are unable to produce myelin.

The region of the neuronal cell body where the axon originates has several specialized features. This domain, called the axon hillock, is distinguished most readily by a deficiency of Nissl substance. Therefore, protein synthesis cannot take place to any appreciable degree in this region. Cytoplasm in the vicinity of the axon hillock may have a few polysomes but is dominated by the cytoskeletal and membranous organelles that are being delivered to the axon. Microtubules and neurofilaments begin to align roughly parallel to each other, helping to organize membrane-limited organelles destined for the axon. The hillock is a region where materials either are committed to the axon (cytoskeletal elements, synaptic vesicle precursors, mitochondria, etc.) or are excluded from the axon (RER and free polysomes, dendritic microtubule-associated proteins). The molecular basis for this sorting is not understood, although recent studies suggest that actin and microtubule-based cytoskeletal elements may play a role in this sorting at the protein level (Arnold, 2009). Cytoplasm in the axon hillock does not appear to contain a physical "sizing" barrier (like a filter) because large organelles such as mitochondria enter the axon readily, whereas only a small number of essentially excluded structures such as polysomes are occasionally seen only in the initial segment of the axon and not in the axon proper. A known exception to this general rule is during development when local protein synthesis does take place at the axon terminus or growth cone. Nevertheless, new methods of detecting gene transcripts, including so-called "deep sequencing," are beginning to suggest that even in mature axons mRNAs may be more abundant than originally thought, in spite of the fact that classical ribosomes or even polysomes are difficult to detect. In the mature neuron,

FIGURE 4.1 Basic elements of neuronal subcellular organization. The neuron consists of a soma, or cell body, in which the nucleus, multiple cytoplasm-filled processes termed dendrites, and the (usually single) axon are placed. The neuron is highly extended in space; a neuron with a cell body of the size shown here could easily maintain an axon several miles in length! The unique shape of each neuron is the result of a cooperative interplay between plasma membrane components (the lipid matrix and associated proteins) and cytoskeletal elements. Most large neurons in vertebrates are myelinated by oligodendrocytes in the CNS and by Schwann cells in the PNS. The compact wraps of myelin encasing the axon distal to the initial segment permit the rapid conduction of the action potential by a process termed "saltatory conduction" (see Chapter 3).


Large PNS neuron with myelin sheath
DENDRITES
SOMA
Nucleus
Spines
Golgi apparatus
Nissl substance
Axon hillock
Initial segment
Myelin sheath
Node of ranvier
Schwann cell
AXON
Neurofilament
Microtubule
Presynaptic terminal
Synaptic vesicles

the physiological significance of this barrier must be considerable because axonal structures are found to accumulate in this region in many neuropathologies, including those due to degenerative diseases (such as amyotrophic lateral sclerosis) and to exposure to neurotoxic compounds (such as acrylamide).

The initial segment of the axon is the region of the axon adjacent to the axon hillock (Palay, Sotelo, Peters & Orkand, 1968; Grubb & Burrone, 2010). Microtubules generally form characteristic fascicles, or bundles, in the initial segment of the axon. These fascicles are not seen elsewhere in the neuron. The initial segment and, to some extent, the axon hillock also have a distinctive specialized plasma membrane. Initially, the plasmalemma was thought to have a thick electrondense coating actually attached to the inner surface of the membrane, but this dense undercoating is in reality separated by 5–10 nm from the plasma membrane inner surface and has a complex ultrastructure. Neither the composition nor the function of this undercoating is known. Curiously, the undercoating is present in the same regions of the initial segment as the distinctive fasciculation of microtubules, although the relationship is not understood. The plasma membrane is specialized in the initial segment and axon hillock in that it contains voltage-sensitive ion channels in large numbers together with several other proteins known to play a part in either clustering the voltage-gated sodium channels (ankyrin G), or stabilizing them (Neurofascin 186). It has generally been thought that most action potentials are initiated in this domain, but it may also have a role in modulating action potentials that arise elsewhere.

Ultimately, axonal structure is geared toward the efficient conduction of action potentials at a rate appropriate to the function of that neuron. This can be seen from both the ultrastructure and the composition of axons. Axons are roughly cylindrical in cross section with little or no taper. As discussed later, this diameter is maintained by regulation of the cytoskeleton. Even at branch points, daughter axons are comparable in diameter to the parent axon. This constant caliber helps ensure a consistent rate of conduction. Similarly, the organization of membrane components is regulated to this end. Voltage-gated ion channels are distributed to maximize conduction. Sodium channels are distributed more or less uniformly in small nonmyelinated axons, but are concentrated at high density in the regularly spaced unmyelinated gaps, known as nodes of Ranvier (Sherman & Brophy, 2005; Susuki & Rasband, 2008). The nodes are flanked by specialized domains known as the paranodal axoglial junctions, which, as their name suggests, are sites where junctional complexes form between the terminal loops of the myelin sheath and the axolemma. These play an important role in both clustering sodium channels at the node and restricting them there in mature nerve fibers. Disruption of the septate junctions between the paranodal loops of myelin and the axonal membrane impair rapid nerve impulse conduction. An axon so organized will conduct an action potential or train of spikes long distances with high fidelity at a defined speed. These characteristics are essential for maintaining the precise timing and coordination seen in neuronal circuits.

Most vertebrate neurons have multiple dendrites arising from their perikarya. Unlike axons, dendrites branch continuously and taper extensively with a reduction in caliber in daughter processes at each branching. In addition, the surface of dendrites is covered with small protrusions, or spines, that are postsynaptic specializations. Although the surface area of a dendritic arbor may be quite extensive, dendrites in general remain in the relative vicinity of the perikaryon. A dendritic arbor may be contacted by the axons of many different and distant neurons or innervated by a single axon making multiple synaptic contacts.

The base of a dendrite is continuous with the cytoplasm of the cell body. In contrast to the axon, Nissl substance extends into dendrites, and certain proteins are synthesized predominantly in dendrites (Martin & Zukin, 2006). There is evidence for the selective placement of some mRNAs in dendrites as well. For example, whereas RER and polysomes extend well into the dendrites, the mRNAs that are transported and translated in dendrites are a subset of the total neuronal mRNA, deficient in some mRNA species (such as neurofilament mRNAs) and enriched in mRNAs with dendritic functions (such as microtubule-associated protein, MAP2, mRNAs). Also, certain proteins appear to be targeted, postsynthesis, to the dendritic compartment as well.

The shapes and complexity of dendritic arborizations may be remarkably plastic. Dendrites appear relatively late in development and initially have only limited numbers of branches and spines. As development and maturation of the nervous system proceed, the size and number of branches increase. The number of spines increases dramatically, and their distribution may change. This remodeling of synaptic connectivity may continue into adulthood, and environmental effects can alter this pattern significantly. Eventually, in the aging brain, there is a reduction in complexity and size of dendritic arbors, with fewer spines and thinner dendritic shafts. These changes correlate with changes in neuronal function during development and aging.

As defined by classical physiology, axons are structural correlates for neuronal output, and dendrites constitute the domain for receiving information. The functional and morphological hallmarks of axons and dendrites are listed in Table 4.1. A neuron without an axon or one without dendrites might therefore seem paradoxical, but such neurons do exist. Certain amacrine and horizontal cells in the vertebrate retina have no identifiable axons, although they do have dendritic processes

TABLE 4.1 Functional and Morphological Hallmarks of Axons and Dendrites[a]

Axons	Dendrites
With rare exceptions, each neuron has a single axon	Most neurons have multiple dendrites arising from their cell bodies
Axons appear first during neuronal differentiation	Dendrites begin to differentiate only after the axon has formed
Axon initial segments are distinguished by a specialized plasma membrane containing a high density of ion channels and distinctive cytoskeletal organization	Dendrites are continuous with the perikaryal cytoplasm, and the transition point cannot be distinguished readily
Axons typically are cylindrical in form with a round or elliptical cross section	Dendrites usually have a significant taper and small spinous processes that give them an irregular cross section
Large axons are myelinated in vertebrates, and the thickness of the myelin sheath is proportional to the axonal caliber	Dendrites are not myelinated, although a few wraps of myelin may occur rarely
Axon caliber is a function of neurofilament and microtubule numbers with neurofilaments predominating in large axons	The dendritic cytoskeleton may appear less organized, and microtubules dominate even in large dendrites
Microtubules in axons have a uniform polarity with plus ends distal from the cell body	Microtubules in proximal dendrites have mixed polarity, with both plus and minus ends oriented distal to the cell body
Axonal microtubules are enriched in tau protein with a characteristic phosphorylation pattern	Dendritic microtubules may contain some tau protein, but MAP2 is not present in axonal compartments and is highly enriched in dendrites
Ribosomes are excluded from mature axons, although a few may be detectable in initial segments	Both rough endoplasmic reticulum and cytoplasmic polysomes are present in dendrites, with specific mRNAs being enriched in dendrites
Axonal branches tend to be distal from the cell body	Dendrites begin to branch extensively near the perikaryon and form extensive arbors in the vicinity of the perikaryon
Axonal branches form obtuse angles and have diameters similar to the parent stem	Dendritic branches form acute angles and are smaller than the parent stem
Most axons have presynaptic specializations that may be *en passant* or at the ends of axonal branches	Dendrites are rich in postsynaptic specializations, particularly on the spinous processes that project from the dendritic shaft
Action potentials are usually generated at the axon hillock and conducted away from the cell body	Some dendrites can generate action potentials, but more commonly they modulate the electrical state of the perikaryon and initial segment
Traditionally, axons are specialized for conduction and synaptic transmission—that is, neuronal output	Dendritic architecture is most suitable for integrating synaptic responses from a variety of inputs—that is, neuronal input

[a]*Neurons typically have two classes of cytoplasmic extensions that may be distinguished using electrophysiological, morphological, and biochemical criteria. Although some neuronal processes may lack one or more of these features, enough parameters can generally be defined to allow unambiguous identification.*

that are morphologically distinct from axons. Such dendritic processes may have both pre- and postsynaptic specializations or may have gap junctions that act as direct electrical connections between two cells. Similarly, the pseudounipolar sensory neurons of dorsal root ganglia (DRG) have no dendrites. In their mature form, these DRG sensory neurons give rise to a single axon that extends a few hundred micrometers before branching. One long branch extends to the periphery, where it may form a sensory nerve ending in muscle spindles or skin. Large DRG peripheral branches are myelinated and have the morphological characteristics of an axon, but they contain neither pre- nor postsynaptic specializations. The other branch extends into the central nervous system, where it forms synaptic contacts. In DRG neurons, the action potential is generated at distal sensory nerve endings and is then transmitted along the peripheral branch to the central branch and the appropriate central nervous system (CNS) targets, bypassing the cell body.

Summary

Neurons are polarized cells that are specialized for membrane and protein synthesis, as well as for conduction of the nerve impulse. In general, neurons have a cell body, a dendritic arborization that is usually located near the cell body, and an extended axon that may branch considerably before terminating to form synapses with other neurons.

PROTEIN SYNTHESIS IN NERVOUS TISSUE

Both neurons and glial cells have strikingly extended morphologies. This cytoarchitecture is ideal for a tissue whose functions depend on multiple intercellular contacts locally and at great distances. Protein and lipid components are synthesized and assembled into the membranes of these cell extensions through pathways of membrane biogenesis that have been elucidated primarily in other cell types, including the yeast *Saccharomyces cerevisiae*. However, some adaptations of these general mechanisms have been necessary, due to the specific requirements of cells in the nervous system. Neurons, for example, have devised mechanisms for ensuring that the specific components of the axonal and dendritic plasma membranes are selectively delivered (targeted) to each plasma membrane subdomain.

The distribution to specific loci of organelles, receptors, and ion channels is critical to normal neuronal function. In turn, these loci must be "matched" appropriately to the local microenvironment and specific cell–cell interactions. Similarly, in myelinating glial cells during the narrow developmental window when the myelin sheath is being formed, these cells synthesize vast sheets of insulating plasma membrane at an unbelievably high rate. To understand how the plasma membrane of neurons and glia might be modeled to fit individual functional requirements, it is necessary to review the progress that has been made so far in our understanding of how membrane components and organelles are generated in eukaryotic cells.

There are two major categories of membrane proteins: integral and peripheral. Integral membrane proteins, which include receptors for neurotransmitters (e.g., acetylcholine receptor subunits) and polypeptide growth factors (e.g., the dimeric insulin receptor) as well as ion channels, have segments that are either embedded in the lipid bilayer or bound covalently to molecules that insert into the membrane, such as those proteins linked to glycosyl phosphatidylinositol at their C termini (e.g., Thy–1). A protein with a single membrane-embedded segment and an N terminus exposed at the extracellular surface is said to be of type I, whereas type II proteins retain their N termini on the cytoplasmic side of the plasma membrane. Peripheral membrane proteins are localized on the cytoplasmic surface of the membrane and do not cross any membrane during their biogenesis. They interact with the membrane either by means of their associations with membrane lipids or the cytoplasmic tails of integral proteins or by means of their affinity for other peripheral proteins (e.g., platelet-derived growth factor receptor-Grb2-Sos-Ras complex). In some cases, they may bind directly to the polar head groups of the lipid bilayer (e.g., myelin basic protein).

Integral Membrane and Secretory Polypeptides Are Synthesized *de Novo* in the Rough Endoplasmic Reticulum

Subcellular destinations for integral and peripheral membrane proteins may be determined by their sites of synthesis. In the secretory pathway, integral membrane proteins and secretory proteins are synthesized in the rough endoplasmic reticulum, whereas the mRNAs encoding peripheral proteins are translated on cytoplasmic "free" polysomes, which are not membrane associated but may interact with cytoskeletal structures.

The pathway by which secretory proteins are synthesized and exported was first postulated through the elegant ultrastructural studies on the pancreas by George Palade and colleagues (Palade, 1975). Pancreatic acinar cells were an excellent choice for this work because they are extremely active in secretion, as revealed by the abundance of their RER network, a property they share with neurons. Nissl deduced, in the nineteenth century, that pancreatic cells and neurons would be found to have common secretory properties (Fig. 4.2), because of similarities in the distribution of the Nissl substance.

Pulse–chase autoradiography revealed that newly synthesized secretory proteins move from the RER to the Golgi apparatus, where the proteins are packaged

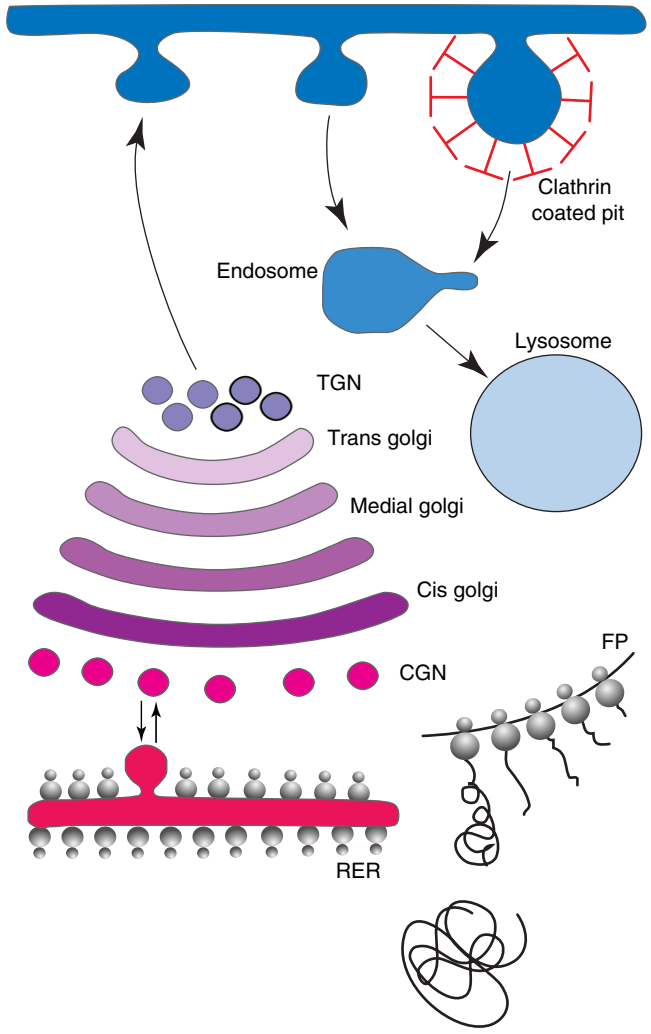


FIGURE 4.2 The secretory pathway. Transport and sorting of proteins in the secretory pathway occur as they pass through the Golgi complex before reaching the plasma membrane. Sorting occurs in the *cis*-Golgi network (CGN), also known as the intermediate compartment, and in the *trans*-Golgi network (TGN). Proteins exit from the Golgi complex at the TGN. The default pathway is the direct route to the plasma membrane. Proteins bound for regulated secretion or for transport to endosomes and from there to lysosomes are diverted from the default path by means of specific signals. In endocytosis, one population of vesicles is surrounded by a clathrin cage and is destined for late endosomes.

into secretory granules and transported to the plasma membrane, from which they are released by exocytosis. Pulse–chase studies in neurons reveal a similar sequence of events for proteins transported into the axon (Brady, 1985). Unraveling the detailed molecular mechanisms of the pathway began with the successful reconstitution of secretory protein biosynthesis *in vitro* and the direct demonstration that, very early during synthesis, secretory proteins are translocated into the lumen of RER vesicles, prepared by cell fractionation, termed microsomes. A key observation here was that the fate of the protein was sealed as a result of encapsulation in the lumen of the RER at the site of synthesis. This cotranslational insertion model provided a logical framework for understanding the synthesis of integral membrane proteins with a transmembrane orientation.

The process by which integral membrane proteins are synthesized closely follows the secretory pathway, except that integral proteins are of course not released from the cell, but instead remain within cellular membranes (Rapoport, 2007). Synthesis of integral proteins begins with synthesis of the nascent chain on a polysome that is not yet bound to the RER membrane (Fig. 4.3). Emergence of the N terminus of the nascent protein from the protein synthesizing machinery allows a ribonucleoprotein, a signal recognition particle (SRP), to bind to an emergent hydrophobic signal sequence and prevent further translation (Walter & Johnson, 1994). Translation arrest is relieved when SRP docks with its cognate receptor in the RER and dissociates from the signal sequence in a process that requires GTP. Synthesis of transmembrane proteins on RER is an extremely energy-efficient process. The passage of a fully formed and folded protein through a membrane is, thermodynamically, formidably expensive; it is infinitely "cheaper" for cells to thread amino acids, in tandem, through a membrane during initial protein synthesis.

Protein synthesis then resumes, and the emerging polypeptide chain is translocated into the RER membrane through a conceptualized "aqueous pore" termed the translocon. Several proteins have been identified in cross-linking experiments as possible components of the translocon, including translocating chain-associating membrane protein (TRAM) and translocon-associated protein a (TRAP), which is a component of the mammalian homologue of the yeast sec 61 complex. Many others are strikingly similar to proteins that were originally discovered in yeast, revealing the common conserved nature of this process in organisms as diverse as yeasts and humans.

A few polypeptides deviate from the common pathway for secretion. For example, certain peptide growth factors, such as basic fibroblast growth factor and ciliary neurotrophic factor, are synthesized without signal peptide sequences but are potent biological modulators of

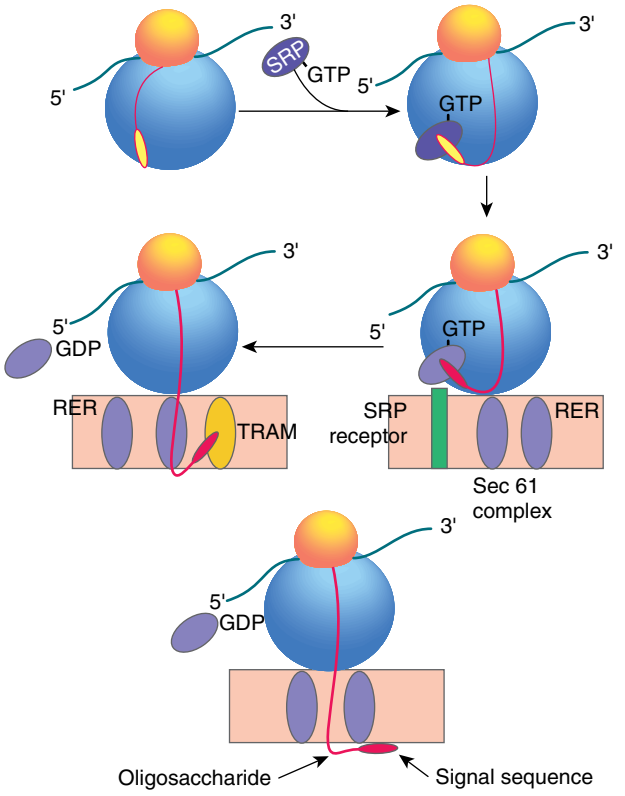


FIGURE 4.3 Translocation of proteins across the rough endoplasmic reticulum (RER). Integral membrane and secretory protein synthesis begins with partial synthesis on a free polysome not yet bound to the RER. The N terminus of the nascent protein emerges and allows a ribonucleoprotein, signal recognition particle (SRP), to bind to the hydrophobic signal sequence and prevent further translation. Translation arrest is relieved once the SRP docks with its receptor at the RER and dissociates from the signal sequence in a process that requires GTP. Once protein synthesis resumes, translocation occurs through an aqueous pore termed the translocon, which includes translocating chain-associating membrane protein (TRAM) and translocon-associated protein (TRAP). The signal sequence is removed by a signal peptidase located in the lumen of the RER.

cell survival and differentiation. These growth factors appear to be released under certain conditions, although the mechanisms for such release are still controversial. One possibility is that release of these factors may be associated primarily with cellular injury.

Two cotranslational modifications are commonly associated with the emergence of the polypeptide on the luminal face of the RER (Fig. 4.3). First, an N-terminal hydrophobic signal sequence that is used for insertion into the RER is usually removed by a signal peptidase. Second, oligosaccharides rich in mannose sugars are transferred from a lipid carrier, dolichol phosphate, to the side chains of asparagine residues (Kornfeld & Kornfeld, 1985). The asparagines must be in the sequence N X T (or S), and they are linked to mannose sugars by two molecules of *N*-acetylglucosamine. Although the prevention of glycosylation of some proteins causes their aggregation and

accumulation in the RER and Golgi apparatus, for most glycoproteins, the significance of glycosylation is not apparent. Neither is it a universal feature of integral membrane proteins: some proteins, such as the proteolipid proteins of CNS myelin, neither lose their signal sequence nor become glycosylated, although they do appear to be posttranslationally modified by the addition of the fatty acid, palmitic acid. One clear case of a proved function for a carbohydrate moiety is the targeting of proteins to lysosomes in the *trans*-Golgi by means of the mannose 6-phosphate receptor (Braulke & Bonifacino, 2009). Intercellular adhesion molecules, such as selectins, which affect the sticking of lymphocytes to blood vessel walls, appear to interact with lectin-like proteins through their oligosaccharide chains. Similarly, the sialic acid side chains of the neural cell adhesion molecule (NCAM) are essential for modulation of cell–cell adhesion mediated by NCAMs. Thus, for the vast majority of polypeptides destined for release from the cell (secretory polypeptides), an N-terminal "signal sequence" first mediates the passage of the protein into the RER and is cleaved immediately from the polypeptide by a signal peptidase residing on the luminal side of the RER. For proteins destined to remain as permanent residents of cellular membranes (and these form a particularly important and diverse category of plasma membrane proteins in neurons and myelinating glial cells), however, many variations on this basic theme have been found. Simply stated:

1. Signal sequences for membrane insertions need not be only N-terminal; those that lie within a polypeptide sequence are not cleaved.
2. A second type of signal, a "halt" or "stop" transfer signal, functions to arrest translocation through the membrane bilayer. The halt transfer signal is also hydrophobic and is usually flanked by positive charges. This arrangement effectively stabilizes a polypeptide segment in the RER membrane bilayer.
3. The sequential display in tandem of insertion and halt transfer signals in a polypeptide as it is being synthesized ultimately determines its disposition with respect to the phospholipid bilayer, and thus its final topology in its target membrane. By synthesizing transmembrane polypeptides in this way, virtually any topology may be generated.

Newly Synthesized Polypeptides Exit from the RER and Are Moved through the Golgi Apparatus

When the newly synthesized protein has established its correct transmembrane orientation in the RER, it is incorporated into vesicles and must pass through the Golgi complex before reaching the plasma membrane (Fig. 4.2). For membrane proteins, the Golgi serves two major functions: it sorts and targets proteins, and it performs further posttranslational modifications, particularly on the oligosaccharide chains that were added in the RER (Glick & Nakano, 2009). Sorting takes place in the *cis*-Golgi network (CGN), also known as the intermediate compartment, and in the *trans*-Golgi network (TGN), whereas sculpting of oligosaccharides is primarily the responsibility of the *cis*-, *medial*-, and *trans*-Golgi stacks. The TGN is a tubulovesicular network wherein proteins are targeted to the plasma membrane or to organelles.

In addition to the processing of carbohydrates in the Golgi, posttranslational modifications can take place in other subcellular compartments. Some protein glycosylations are modified further post-Golgi in components of the smooth endoplasmic reticulum or transport vesicles, as described later in this section. Finally, some neuropeptides (vasopressin, enkephalins, etc.) are synthesized as sequence domains in large precursor proteins that must be cleaved in transit by specific proteases to form the biologically active form.

The CGN serves an important sorting function for proteins entering the Golgi from the RER. Because most proteins that move from the RER through the secretory pathway do so by default, any resident endoplasmic reticulum proteins must be restrained from exiting or returned promptly to the RER from the CGN should they escape (Dudek et al., 2009). Although no retention signal has been demonstrated for the endoplasmic reticulum, two retrieval signals have been identified: a Lys-Asp-Glu-Leu or KDEL sequence in type I proteins and the Arg-Arg or RR motif in the first five amino acids of proteins with a type II orientation in the membrane. The KDEL tetrapeptide binds to a receptor called Erd 2 in the CGN, and the receptor–ligand complex is returned to the RER. There may also be a receptor for the N arginine dipeptide; alternatively, this sequence may interact with other components of the retrograde transport machinery, such as microtubules.

Movement of proteins between Golgi stacks proceeds by means of vesicular budding and fusion (Glick & Nakano, 2009). Through the use of a cell-free assay containing Golgi-derived vesicles, the essential mechanisms for budding and fusion have been shown to require coat proteins (COPs) in a manner that is analogous to the role of clathrin in endocytosis. Currently, two main types of COP complex, COPI and COPII, have been distinguished. Although both have been shown to coat vesicles that bud from the endoplasmic reticulum, they may have different roles in membrane trafficking. Coat proteins provide the external frame-work into which a region of a flattened Golgi cisternae can bud and vesiculate. A complex of these COPs forms the coatomer (coat protomer) together with a p200 protein, and a family of

GTP-binding proteins called ADP-ribosylation factors (ARFs), originally named for their role in the action of cholera toxin. Immunolocalization of one of the coatamer proteins, β-COP, predominantly to the CGN and *cis*-Golgi indicates that these proteins may also take part in vesicle transport into the Golgi (Fig. 4.4A). The function of ARF is to drive the assembly of the coatamer and therefore vesicle budding in a GTP-dependent fashion. Dissociation of the coat is triggered when a GTPase-activating protein (GAP) in the Golgi membrane stimulates hydrolysis of the GTP bound to ARF. The cycle of coat assembly and disassembly can continue when the replacement of GDP on ARF by GTP is catalyzed by a guanine nucleotide exchange factor (GEF). The importance of this GDP–GTP exchange to normal vesicular traffic is illustrated dramatically by the effects of brefeldin A, a fungal metabolite that specifically inhibits GTP exchange and disperses the Golgi complex by preventing the return of Golgi components from the intermediate compartment.

Fusion of vesicles with their target membrane in the Golgi apparatus is regulated by a series of proteins, *N*-ethylmaleimide-sensitive factor (NSF), soluble NSF attachment proteins (SNAPs), and SNAP receptors (SNAREs), which together assist the vesicle in docking with its target membrane. In addition, Rabs, a family of membrane-bound GTPases, act in concert with their own GAPs, GEFs, and a cytosolic protein that dissociates Rab–GDP from membranes after fusion called guanine-nucleotide dissociation inhibitor. Rabs are believed to regulate the action of SNAREs, the proteins directly engaged in membrane–membrane contact prior to fusion. The tight control necessary for this process and the importance of ensuring that vesicle fusion takes place only at the appropriate target membrane may explain why eukaryotic cells contain so many Rabs, some of which are known to specifically take part in the internalization of endocytic vesicles at the plasma membrane (Fig. 4.2).

Exocytosis of the neurotransmitter at the synapse must occur in an even more finely regulated manner than endocytosis (Sudhof, 2004). The proteins first identified in vesicular fusion events in the secretory pathway (namely NSF, SNAPs, and SNAREs or closely related homologues) appear to play a part in the fusion of synaptic vesicles with the active zones of the presynaptic neuronal membrane (Fig. 4.4B).

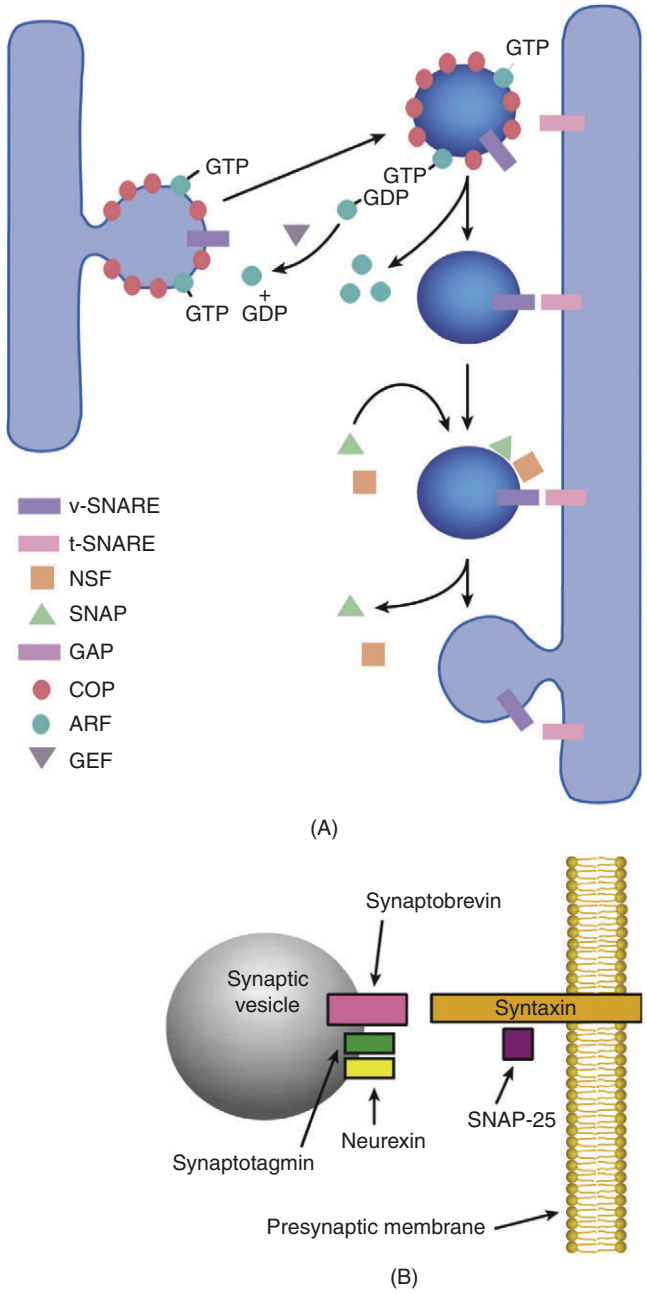


FIGURE 4.4 **A.** General mechanisms of vesicle targeting and docking in the ER and Golgi. The assembly of coat proteins (COPs) around budding vesicles is driven by ADP-ribosylation factors (ARFs) in a GTP-dependent fashion. Dissociation of the coat is triggered when hydrolysis of the GTP bound to ARF is stimulated by a GTPase activating protein (GAP) in the Golgi membrane. The cycle of coat assembly and disassembly can continue when the replacement of GDP on ARF by GTP is catalyzed by a guaninenucleotide exchange factor (GEF). Fusion of vesicles with their target membrane in the Golgi is regulated by a series of proteins, *N*-ethyl-maleimide-sensitive factor (NSF), soluble NSF attachment proteins (SNAPs), and SNAP receptors (SNAREs), which together assist the vesicle in docking with its target membrane. SNAREs on the vesicle (v-SNAREs) are believed to associate with corresponding t-SNAREs on the target membrane. **B.** Mechanisms of vesicle targeting and docking in the synaptic terminal. The synaptic counterpart of v-SNARE is synap-tobrevin (also known as vesicle-associated membrane protein), and syntaxin corresponds to t-SNARE. SNAP-25 is an accessory protein that binds to syntaxin. Synaptotagmin is believed to be the Ca^{2+} sensitive regulatory protein in the complex that binds to syntaxin. Neurexins appear to have a role in conferring Ca^{2+} sensitivity to these interactions.

The synaptic counterpart of vSNARE is synaptobrevin [also known as vesicle-associated membrane protein (VAMP)], and syntaxin corresponds to t-SNARE. SNAP-25 is an accessory protein that binds to syntaxin. In the constitutive pathway, such as between the RER and Golgi apparatus, assembly of the complex at the target membrane promotes fusion. However, at the presynaptic membrane, Ca^{2+} influx is required to stimulate membrane fusion at the presynaptic membrane. Synaptotagmin is believed to be the Ca^{2+}-sensitive regulatory protein in the complex that binds syntaxin. Neurexins appear to have a role in regulation as well, because, in addition to interacting with synaptotagmin, they are the targets of black widow spider venom (α)-latrotoxin, which deregulates the Ca^{2+}-dependent exocytosis of the neurotransmitter.

The comparison between secretion in slow-releasing cells, such as the pancreatic β-cell, and neurotransmitter release at the neuromuscular junction is much like the comparison between a hand-held pocket calculator for balancing a checkbook and a state-of-the-art desktop computer. Two differences stand out. First, the speed of neurotransmitter release is much greater both in release from a single vesicle and in total release in response to a specific signal. Releasing the contents of a single synaptic vesicle at a mouse neuromuscular junction takes from 1 to 2 ms, and the response to an action potential involving the release of many synaptic vesicles is over in approximately 5 ms. In contrast, releasing the insulin in a single secretory granule by a pancreatic β-cell takes from 1 to 5 s, and the full release response may take from 1 to 5 min. A 103- to 105-fold difference in rate is an extraordinary range, making neurotransmitter release one of the fastest biological events routinely encountered, but this speed is critical for a properly functioning nervous system.

A second major difference between slow secretion and fast secretion is seen in the recycling of vesicles. In the pancreas, secretory vesicles carrying insulin are used only once, and so new secretory vesicles must be assembled *de novo* and released from the TGN to meet future requirements. In the neuron, the problem is that the synapse may be at a distance of 1 m or more from the protein synthetic machinery of the perikaryon, and so newly assembled vesicles even traveling at rapid axonal transport rates (see later) may take more than a day to arrive. Now, the number of synaptic vesicles released in 15 min of constant stimulation at a single frog neuromuscular junction has been calculated to be on the order of 105 vesicles, but a single terminal may have only a few hundred vesicles at any one time. These measurements would make no sense if synaptic vesicles had to be replaced constantly through new synthesis in the perikaryon, as is the case with insulin-carrying vesicles. The reason that these numbers are possible is that synaptic vesicles are taken up locally by endocytosis, refilled with neurotransmitter, and reutilized at a rate fast enough to keep up with normal physiological stimulation levels. This takes place within the presynaptic terminal, and evidence shows that these recycled synaptic vesicles are used preferentially (Sudhof, 2004). Such recycling does not require protein synthesis because the classical neurotransmitters are small molecules, such as acetylcholine, or amino acids, such as glutamate, that can be synthesized or obtained locally. Significantly, neurons have fast and slow secretory pathways operating in parallel in the presynaptic terminal (Sudhof, 2004). Synapses that release classical neurotransmitters (acetylcholine, glutamate, etc.) with these fast kinetics also contain dense core granules containing neuropeptides (calcitonin gene related peptide, substance P, etc.) that are comparable to the secretory granules of the pancreatic β-cell. These are used only once because neuropeptides are produced from large polypeptide precursors that must be made by protein synthesis in the cell body. The release of neuropeptides is relatively slow; as is the case in endocrine release, neuropeptides serve primarily as modulators of synaptic function. The small, clear synaptic vesicles containing the classic neurotransmitters can in fact be depleted pharmacologically from the presynaptic terminal, while the dense core granules remain. These observations indicate that even though fast and slow secretory mechanisms have many similarities and may even have common components, in neurons they can operate independent of one another (Xu & Xu, 2008).

Proteins Exit the Golgi Complex at the *trans*-Golgi Network

Transport of cargoes in vesicles beyond the TGN is believed to involve coat proteins other than the classical COPs. These are the adaptins, which form Adaptor Protein (AP) complexes. The APs were until recently considered to comprise four families (AP-1-4) where AP-1 and AP-2 segregated their cargoes to clathrin-coated vesicles, whereas AP-3 and AP-4 did not appear to require clathrin. In general it is believed that AP-1 and AP-4 are involved in moving vesicles between the TGN and endosomes, AP-2 is specialized for endocytosis, and AP-3 ferries cargoes from endosomes to lysosomes. Interestingly, mutations in AP-4 cause cerebral palsy, intellectual disability, and spastic paraplegia. Recently a fifth AP complex, AP-5, has been identified that also appears to be involved in endosomal trafficking. Defects in AP-5 have been linked to spastic paraplegia (Hirst et al., 2011).

Most of the N-linked oligosaccharide chains acquired at the RER are remodeled in the Golgi cisternae, and while the proteins are in transit, another type of glycosyl

linkage to serine or threonine residues through *N*-acetylgalactosamine can also be made. Modification of existing sugar chains by a series of glycosidases and the addition of further sugars by glycosyl transferases occur from the *cis* to the *trans* stacks. Some of these enzymes have been localized to particular cisternae. For example, the enzymes (β)-1,4-galactosyltransferase and (α)-2,6-sialyltransferase are concentrated in the *trans*-Golgi. How they are retained there is a matter of some debate. One idea is that these proteins are anchored by oligomerization. Another view is that the progressively rising concentration of cholesterol in membranes more distal to the ER in the secretory pathway increases membrane thickness, which in turn anchors certain proteins and causes an arrest in their flow along the default route.

The default or constitutive pathway seems to be the direct route to the plasma membrane taken by vesicles that bud from the TGN (Fig. 4.2). This is how, in general, integral plasma membrane proteins reach the cell surface. Proteins bound for regulated secretion or for transport to endosomes and from there to lysosomes are diverted from the default path by means of specific signals. It has been assumed that the sorting of proteins for their eventual destination takes place at the TGN itself. However, recent analyses of the three-dimensional structure of the TGN have provoked a revision of this view. These studies have shown that the TGN is tubular, with two major types of vesicles that bud from distinct populations of tubules. The implication is that sorting may already have occurred in the *trans*-Golgi prior to the proteins' arrival at the TGN. One population of vesicles consists of those surrounded by the familiar clathrin cage, which are destined for late endosomes. The β-COP protein and related coatomer proteins active in more proximal regions of the secretory pathway are absent from the TGN.

Endocytosis and Membrane Cycling Occurs in the *trans*-Golgi Network

Two types of membrane invagination occur at the surface of mammalian cells and are clearly distinguishable by electron microscopy (Doherty & McMahon, 2009). The first type is a caveola, which has a thread-like structure on its surface made of the protein caveolin. Caveolae mediate the uptake of small molecules such as vitamin folic acid by a process called potocytosis. They may also have a role in concentrating proteins linked to the plasma membrane by the glycosylphosphatidylinositol anchor. Demonstration of the targeting of protein tyrosine kinases to caveolae by the tripeptide signal MGC (Met-Gly-Cys) also suggests that caveolae may function in signal transduction cascades.

The other type of endocytic vesicle at the cell surface is that coated with the distinctive meshwork of clathrin triskelions (Doherty & McMahon, 2009). The triskelion comprises three copies of a clathrin heavy chain and three copies of a clathrin light chain. The ease with which these triskelions can assemble into a cage structure demonstrates how they promote the budding of a vesicle from a membrane invagination. Clathrin binds selectively to regions of the cytoplasmic surface of membranes that are selected by adaptins. The AP-2 complex, which is primarily active at the plasma membrane, consists of 100-kDa α and β subunits and two subunits of 50 and 17 kDa each. AP-1 complexes localize to the TGN and have γ and subunits of 100 kDa together with smaller polypeptides of 46 and 19 kDa. Adaptins bind to the cytoplasmic tails of membrane proteins, thus recruiting clathrin for budding at these sites.

A further component of the endocytic complex at the plasma membrane is GTPase dynamin, which seems to be required for the normal budding of coated vesicles during endocytosis. Dynamins are a family of 100-kDa GTPases found in both neuronal and nonneuronal cells and may interact with the AP-2 component of a clathrin-coated pit (Doherty & McMahon, 2009). Dynamin I is found primarily in neurons, whereas dynamin II has a widespread distribution. Oligomers of dynamin form a ring at the neck of a budding clathrin-coated vesicle, and GTP hydrolysis appears to be necessary for the coated vesicle to pinch off from the plasma membrane. The existence of a specific neuronal form of dynamin may be a manifestation of the unusually rapid rate of synaptic vesicle recycling.

The primary function of clathrin-coated vesicles at the plasma membrane is to deliver membrane proteins together with any ligands bound to them to the early endosomal apparatus. The other major site of action of clathrin is in vesicles that bud from the TGN carrying lysosomal enzymes en route to late endosomes. Early endosomes have a tubulovesicular morphology. Receptors that will be recycled back to the plasma membrane partition into the tubules; those endocytosed proteins destined for lysosomes concentrate in the vesicular regions. Recycling seems to be the default pathway, whereas proteins must be actively targeted to lysosomes. However, the precise signals for this are unknown.

Regulation of membrane cycling in the endosomal compartment is likely to include the Rab family of small GTP-binding proteins (Zerial & McBride, 2001). Indeed, each stage of the endocytic pathway may have its own Rab protein to ensure efficient targeting of the vesicle to the appropriate membrane. Rab6 is believed to have a role in transport from the TGN to endosomes, whereas Rab9 may regulate vesicular flow in the reverse direction. In neurons, Rab5a has a role in regulating the fusion of endocytic vesicles and early endosomes and appears to function in endocytosis from both somatodendritic domains and the axon. The association of the

protein with synaptic vesicles in nerve terminals, attached presumably by means of its isoprenoid tail, also suggests that early endosomal compartments may have a role in the packaging and recycling of synaptic vesicles.

Lysosome Is the Target Organelle in Several Inherited Diseases that Affect the Nervous System

Lysosomes were first isolated and characterized as a distinct organelle fraction bounded by a single membrane and separable from mitochondria by differential and sucrose gradient centrifugation. Because of their high content of acid hydrolases, the classic view is that lysosomes are organelles of terminal degradation. Indeed, the latency of hydrolase activity before membrane permeabilization by agents such as nonionic detergents has been used biochemically as a measure of the purity and intactness of lysosomal preparations. However, in addition to their well-established function in lipid and protein breakdown, tubulovesicular lysosomes may overlap in sorting functions with early endosomes, particularly during antigen processing in macrophages.

Inherited deficiencies in lipid metabolism in the lysosome often have particularly devastating consequences on the nervous system because of the abundance of the lipid-rich membrane myelin. Metachromatic leukodystrophy is an autosomal recessive disease caused by a deficiency in arylsulfatase A activity, which is also responsible for degrading the myelin lipid cerebroside sulfate (sulfatide). Oligodendrocytes accumulate sulfatide in metachromatic granules, causing severe disruption of myelination. Peripheral nerve myelination is also affected, as are other organs that normally contain much lower amounts of sulfatide, such as the kidney, the liver, and the endocrine system. Krabbe disease, or globoid cell leukodystrophy, is also a dysmyelinating disease in which there is an almost complete lack of oligodendrocytes and therefore myelin, caused by a deficiency in the β-galactosidase responsible for hydrolyzing galactocerebroside to ceramide and galactose. Galactocerebroside is particularly abundant in myelin, constituting about 25% of myelin lipid. Mice that lack galactocerebroside have the ability to assemble multilamellar myelin; however, this myelin does not support adequate nerve conduction, and is not stable. Unlike metachromatic leukodystrophy, Krabbe disease is limited to the CNS and peripheral nervous system (PNS). However, why a buildup of galactocerebroside should prove particularly toxic to oligodendrocytes is not entirely clear. One hypothesis is that a metabolite of galactocerebroside, galactosphingosine (psychosine), is the primary culprit.

How are proteins destined to operate in lysosomes targeted to these organelles? Soluble lysosomal hydrolase enzymes acquire a phosphorylated mannose on their oligosaccharide chains by a two-step process in the Golgi apparatus (Braulke & Bonifacino, 2009). This mannose 6-phosphate label is recognized by specific mannose 6-phosphate receptors, which carry the proteins to late endosomes. In contrast, lysosomal membrane proteins are targeted by means of cytoplasmic tail signals that contain either leucine or tyrosine of type LJ or type YXXJ or NXXY, where J is any hydrophobic amino acid. The LJ signal seems to be essential for efficient delivery directly to endosomes, whereas the second type of signal seems to be more important in the recovery of proteins destined for lysosomes from the plasma membrane. The majority of lysosomal membrane proteins have the YXXJ but do not have the LJ signal. The implication of these observations is that many of the lysosomal membrane proteins make their way to lysosomes from the TGN endosomes through the plasma membrane.

What are the receptors for the type LJ or the type YXXJ or NXXY motifs at the TGN? Because transport from the TGN to the endosomes occurs in clathrin-coated vesicles, the proteins that link such vesicles to membranes, the adaptins, may play a role. The weight of the evidence suggests that AP-1 recognizes the LJ sequence, whereas AP-2 identifies the YXXJ and NXXY motifs. Once the ligands are bound, these adaptins would direct transport of their respective ligands to the endosomes from the TGN or through the plasma membrane, respectively. At present, it is not clear how proteins in the late endosome, such as mannose 6-phosphate receptors, that cycle back to the Golgi are sorted from those whose ultimate destination is a lysosome.

How are Peripheral Membrane Proteins Targeted to Their Appropriate Destinations?

Peripheral membrane proteins are synthesized in the same type of free polysome in which the bulk of the cytosolic proteins are made. However, the cell must ensure that these membrane proteins are sent to the plasma membrane rather than allowed to attach in a haphazard way to other intracellular organelles. The fact that complex machinery has evolved to ensure the correct delivery of integral membrane proteins suggests that some equivalent targeting mechanism must exist for proteins that attach to the cytoplasmic surface of the plasma membrane. Such proteins are translated on "free" polysomes, but these polysomes are associated with cytoskeletal structures and are not distributed uniformly throughout the cell body. In a number of cases, mRNAs that encode soluble cytosolic proteins are

concentrated in discrete regions of the cell, resulting in a local accumulation of the translated protein close to the site of action. For some peripheral membrane proteins, this is the plasma membrane.

Evidence that this mechanism might operate in peripheral membrane protein synthesis came from studies showing biochemically and by *in situ* hybridization that mRNAs encoding the myelin basic proteins are concentrated in the myelinating processes that extend from the cell body of oligodendrocytes (Colman, Kreibich, Frey, & Sabatini, 1982). Myelin basic protein may be a special case because of its very strong positive charge and consequent propensity for binding promiscuously to the negatively charged polar head groups of membrane lipids. Nevertheless, the fact that actin mRNAs are localized to the leading edge of cultured myocytes and mRNA for the microtubule-associated protein MAP2b is concentrated in the dendrites of neurons suggest that targeting by local synthesis is more common than originally thought. This mechanism is probably less important for peripheral membrane proteins that associate with the cytoplasmic surface of the plasma membrane by means of strong specific associations with proteins already located at the membrane because such proteins would act as specific receptors. Because only selected cytoplasmic mRNAs are localized to the periphery, the process is specific. However, no mRNAs are localized exclusively to the periphery, and a significant fraction is typically localized proximal to the nucleus in a region rich with translational and protein-processing machinery of the cell (the Nissl substance or translational cytoplasm).

Special Mechanisms Are Used to Target Proteins to Mitochondria and Peroxisomes

The inner membrane of the mitochondrion is the site of oxidative phosphorylation in which the step-by-step transfer of electrons from oxygen intermediary metabolites to molecular oxidation is coupled to proton transport and ATP synthesis. Thus, this organelle has an essential role in providing the large amount of ATP required for the electrical activity of neurons. The fact that, in a resting adult, about 40% of the total energy consumption is required for ion pumping in the CNS accounts for the exquisite sensitivity of the brain to damage from oxygen deprivation. The sensitivity of neurons to interruptions in the provision of ATP by the mitochondrion is also seen in cases of uremia, where a buildup of ammonium ions depletes the Krebs cycle of α-oxoglutaric acid by converting it into glutamate.

Although the mitochondrion has its own circular DNA that encodes some proteins, most mitochondrial proteins are synthesized in the nucleocytoplasmic system (Endo & Yamano, 2009). This poses the problem of how these proteins once made in the cytoplasm gain entry into the mitochondrion. Furthermore, because the mitochondrion has an inner and an outer membrane, some proteins must cross two membranes to gain access to the inner matrix. This group of proteins includes the enzymes of the Krebs cycle and the fatty acid β-oxidation pathway.

Unlike proteins inserted into the RER, mitochondrial proteins can be imported either posttranslationally or cotranslationally with the use of a cleavable amphipathic helical signal sequence usually at the N terminus. At the RER, the signal sequence of a nascent polypeptide chain can be translocated across the membrane because the polypeptide remains small and unfolded due to the arrest of translation caused by a signal recognition particle. However, mitochondrial proteins are typically synthesized on cytoplasmic or free polysomes and must be folded at least partially to prevent degradation. For posttranslational import, mitochondria rely on a group of molecular chaperones to prevent complete folding of the polypeptides. These hsp70 and hsp60 proteins were originally identified because they are upregulated during heat shock. Their role in binding to proteins and maintaining them in specific conformations helps explain why these proteins have an important function in protecting proteins against the stress of elevated temperatures, as well as facilitating the proper folding of newly synthesized polypeptides. In yeast, a second protein, Ydj1p, whose bacterial homologue DnaJ regulates chaperone function has been identified. Ydj1p possesses an isoprenoid tail linked to its C-terminal amino acid, which may serve to anchor the protein to the outer membrane. The third factor that has been implicated is the mitochondrial stimulation factor, which is a heterodimer possessing an ATP-dependent protein "unfoldase" activity. This factor may be more important in cotranslational import where polysomes are known to be associated with the mitochondrial outer membrane.

Most of our current understanding of protein translocation from the mitochondrial outer membrane inward has come from studies on either the fungus *Neurospora crassa* or the yeast *S. cerevisiae*. In both yeast and higher eukaryotes, the partially folded polypeptide targeted for the mitochondrion may be stabilized by a cytoplasmic chaperone that is a member of the hsp70 family, but this interaction is not required for import. However, several proteins in the outer membrane form an essential complex that acts as a receptor and pore for protein translocation. This complex can in turn interact with an inner membrane complex at specialized contact sites that minimize the distance across the two membranes, thereby facilitating the movement of proteins to the inner matrix. Although both pores can function independently, they contact and cooperate when there is

a transmembrane potential across the inner membrane. This accounts for early observations showing that importation of subunits of the F1-ATPase, an inner membrane protein, required an active electron transport chain but did not need ATP synthesis. The mitochondrial import sequence extends through the pore into the inner matrix, where a second member of the hsp70 family binds and facilitates movement into the inner matrix. After proteins have crossed into the inner matrix, they must dissociate from hsp70 in order to fold properly, a process that requires another kind of molecular chaperone, hsp60.

Peroxisomes are so named because they contain oxidases that generate H_2O_2 and the enzyme catalase, which is responsible for detoxifying it. In addition, these organelles contain many other enzymes that take part in lipid, purine, and amino acid metabolism. Peroxisomes are of interest because of the number of inherited diseases associated with defects either in certain enzymes or indeed in the assembly of the organelle itself. Some of these diseases manifest as particularly damaging to the nervous system and include adrenoleukodystrophy (accumulation of very long chain fatty acids due to insufficient lignoceryl-CoA ligase activity caused by inefficient import of the protein) and Refsum disease (buildup of phytanic acid due to defective α-oxidation), both of which cause demyelination.

Like many mitochondrial proteins, peroxisomal proteins are imported posttranslationally (Ma, Agrawal, & Subramani, 2011). Although cytosolic factors are implicated in peroxisomal biogenesis, no peroxisomal chaperones analogous to the hsp70 family have yet been shown to function in protein import. Therefore, unfolding and refolding are assumed not to play a role in the accumulation of proteins inside the peroxisome. Among these cytosolic proteins is presumed to be the receptor for the tripeptide C-terminal import signal SKL (Ser-Lys-Leu) known as peroxisomal targeting signal 1 (PTS1). In addition to C-terminal PTS1, some peroxisomal proteins have a cleavable N-terminal sequence called PTS2, which signals their import. A quite distinct translocation machinery appears to operate for PTS1 and PTS2 proteins. Two possible receptor proteins in the peroxisomal membrane, one of which is the adrenoleukodystrophy protein (ALDP), have been identified. ALDP is a member of a larger family known as ABC ATP-dependent membrane transporters.

A characteristic feature of peroxisomal biogenesis is that it is stimulated by drugs whose detoxification requires peroxisomal activity. It is possible that mature peroxisomes are recruited from a pool of precursor organelles, and there is some evidence for the existence of such a population in rat liver. Although the mature organelle appears to be spherical, electron microscopic evidence suggests a peroxisomal reticulum in which synthesis and protein import may take place. Mature peroxisomes might then arise from this reticulum by a process of budding.

Cytoplasmic Proteins Are also Compartmentalized

Membrane-bound organelles are the most familiar form of compartmentation in cells, but cytoplasmic regions of the cell containing metabolic compartments exist as well. Regions of the neuronal or glial cytoplasm may have highly specialized polypeptide compositions that are important for function. For example, the neuronal phosphoprotein synapsin is highly enriched in presynaptic terminals, where it participates in the localization and targeting of synaptic vesicles. Similarly, calmodulin and the glycolytic enzyme aldolase have been localized in muscle cells to the region of the I-band, where they are thought to facilitate coupling of ATP production to contractility.

As mentioned earlier, cytoplasmic proteins are synthesized on cytoplasmic polysomes, termed "free" polysomes to reflect an absence of underlying ER membrane, even though they may be restricted to specific domains of the cell cytoplasm. This restriction is particularly obvious in the neuronal perikaryon, where both cytoplasmic polysomes and membrane-associated polysomes are concentrated in areas near the nucleus and Golgi complex. In addition, cytoplasmic polysomes containing specific mRNAs may be localized to certain regions of the cell, such as the proximal dendrite (those encoding the microtubule-associated protein MAP-2) and the processes of oligodendrocytes (those encoding myelin basic protein). In contrast, the protein synthetic machinery of the polysome appears to be effectively excluded from the mature axon. Therefore, cytoplasmic polysomes are representative of cytoplasmic compartmentation for proteins and nucleic acids.

In most cases, localized cytoplasmic proteins interact with cytoskeletal structures in the cytoplasm (see next section), but macromolecular complexes that form in order to make a cellular process more efficient or free from error have been described. Evidence exists that glycolytic enzymes of neurons and muscle cells may be organized in a labile complex that facilitates energy metabolism, but the existence of such complexes remains controversial.

Perhaps the best characterized cytoplasmic macromolecular complex is the proteasome, which is a large protein complex (2×10^6 Da, sedimenting as a 20S particle) that contains several distinct enzymatic activities, including catalytic sites for both ubiquitin-dependent and ubiquitin-independent proteolysis (Hochstrasser, 1995). Ubiquitin is a small, highly conserved polypeptide that is added covalently to cytoplasmic proteins

targeted for degradation. The catalytic core of the proteasome is a barrel-shaped structure formed by four heptameric stacked rings, but additional proteins (about 16 polypeptides) may interact with the 20S core to form a larger 26S particle. Because proteasomes constitute the primary cytoplasmic pathway for protein degradation (i.e., nonlysosomal pathways), they serve a number of important physiological functions, including regulation of cell proliferation and processing of antigens for presentation. In the nervous system, however, proteasomes are likely to be most important for homeostasis, allowing turnover of cytoplasmic polypeptides at specific sites so that the elaborate cellular extensions of neurons and glia may be maintained.

Cytoplasmic proteins may also be compartmentalized effectively by posttranslational modification. Two types of modification may be particularly important for this kind of compartmentalization. Local activation of kinases can lead to the phosphorylation of proteins in specific domains of the neuron. For example, the reversible phosphorylation of synapsin in the presynaptic terminal appears to be responsible for the targeting of synaptic vesicles to the terminal and for the mobilization of vesicles during prolonged stimulation. An impressive variety of cytoplasmic protein kinases that may be selectively activated to modify serines or threonines presented in distinctive consensus sequences have been described. Distinct from these serine or threonine kinases, a number of other kinases that specifically modify tyrosines can be found in the brain. In some cases, the tyrosine kinase is linked directly to a membrane-spanning receptor and phosphorylates cytoplasmic proteins in the vicinity of the receptor after activation. Completing the cycle of phosphorylation and dephosphorylation are a number of phosphatases with varying specificities. The properties and physiological roles for kinases and phosphatases are discussed in greater detail later.

A second common posttranslational modification of cytoplasmic proteins is the addition of carbohydrate moieties. Whereas modification of membrane-associated proteins in the Golgi complex proceeds by the addition of complex carbohydrates through N linkages on selected asparagines, glycosylated cytoplasmic proteins have simpler carbohydrates added through O linkages to serine or threonine hydroxyls. This modification was first recognized as a feature of many nuclear proteins and components of the nuclear membrane, but subsequent studies showed that a number of cytoplasmic proteins also have O-linked carbohydrates. Unlike phosphorylation, relatively little is known about the functional significance of cytoplasmic glycosylation. Remarkably, however, serines and threonines subject to O-linked glycosylation would also be good sites for phosphorylation by various kinases as well. This congruence raises the possibility that glycosylation and phosphorylation of some cytoplasmic proteins may serve complementary functions.

Summary

Membrane biogenesis and protein synthesis in neurons and glial cells are accomplished by the same mechanisms that have been worked out in great detail in other cell types. Integral membrane proteins are synthesized in the rough endoplasmic reticulum, and peripheral membrane proteins are products of cytoplasmic-free ribosomes that are found in the cell sap. For transmembrane proteins and secretory polypeptides, synthesis in the RER is followed by transport to the Golgi apparatus, where membranes and proteins are sorted and targeted for delivery to precise intracellular locations. It is likely that the neuron and glial cell have evolved additional highly specialized mechanisms for membrane and protein sorting and targeting because these cells are so greatly extended in space, although these additional mechanisms have yet to be fully described. The basic features of the process of secretion, which includes neurotransmitter delivery to presynaptic terminals, are beginning to be understood as well. The key features of this process are apparently common to all cells, including yeast, although the neuron has developed certain specializations and modifications of the secretory pathway that reflect its unique properties as an excitable cell.

CYTOSKELETONS OF NEURONS AND GLIAL CELLS

The cytoskeleton of eukaryotic cells is an aggregate structure formed by three classes of cytoplasmic structural proteins: microtubules (tubulins), microfilaments (actins), and intermediate filaments. Each of these elements exists concurrently and independently in overlapping cellular domains. Most cell types contain one or more examples of each class of cytoskeletal structure, but there are exceptions. For example, mature mammalian erythrocytes contain no microtubules or intermediate filaments, but they do have highly specialized actin cytoskeletons. Among cells of the nervous system, the oligodendrocyte is unusual in that it contains no cytoplasmic intermediate filaments. Typically, each cell type in the nervous system has a unique complement of cytoskeletal proteins that are important for the differentiated function of that cell type (Pigino, Song, Kirkpatrick, & Brady, 2011).

Although the three classes of cytoskeletal elements interact with each other and with other cellular structures, all three are dynamic structures rather than passive structural elements. Their aggregate properties

form the basis of cell morphologies and plasticity in the nervous tissue. In many cases, biochemical specialization in the cytoskeleton is characteristic of a particular cell type, function, and developmental stage. Each type of cytoskeletal element has unique functions that are essential for a working nervous system.

Microtubules Are an Important Determinant of Cell Architecture

Microtubules are nearly ubiquitous components of the cytoskeleton in eukaryotes. They play key roles in intracellular transport, are a primary determinant of cell morphology, are the structural correlate of the mitotic spindle, and form the functional core of cilia and flagella. Microtubules are very abundant in the nervous system, and tubulin subunits of microtubules may constitute more than 10% of total brain protein. As a result, many fundamental properties of microtubules have been defined by using microtubule protein prepared from brain extracts. At the same time, the microtubule cytoskeleton of the neuron has a variety of biochemical specializations that meet the unique demands imposed by the size and shape of the neuron.

Of the various functions defined for microtubules, intracellular transport and the generation of cellular morphology are the most important roles played by microtubules in cells of the nervous system. In part, this comes from their ability to organize cytoplasmic polarity. Microtubules *in vitro* are dynamic, polar structures with plus and minus ends that correspond to the fast- and slow-growing ends, respectively. In contrast, both stable and labile microtubules can be identified *in vivo*, where they help define both microscopic and macroscopic aspects of intracellular organization in cells. Microtubule organization, stability, and composition in nervous tissue are all highly regulated in the nervous system.

By electron microscopy, microtubules appear as hollow tubes 25 nm in diameter and in axons can be up to hundreds of micrometers in length. High resolution electron micrographs also reveal that the walls of microtubules typically comprise 13 protofilaments formed by a linear arrangement of globular subunits, although microtubules with 12 to 14 protofilaments exist in some tissues and organisms. Globular subunits in the walls of a microtubule are heterodimers of α- and β-tubulin, whereas a variety of microtubule-associated proteins bind to the surface of microtubules.

Neuronal microtubules are remarkable for their genetic and biochemical diversity. Multiple genes exist for both α-and β-tubulins. These genes are expressed differentially according to cell type and developmental stage. Some of these genetic isotypes are expressed ubiquitously, whereas others are only turned on at specific times in development, in specific cell types, or both. Most tubulin genes are expressed in nervous tissue, and some appear to be enriched or specific to neurons. When specific isotypes are prepared in a pure form, they show variability in assembly kinetics and ability to bind ligands. However, when more than one isotype is expressed in a single cell, such as a neuron, they coassemble into microtubules with mixed composition.

A variety of posttranslational modifications of tubulins have been described, the most common of which are tyrosination–detyrosination, acetylation–deacetylation, and phosphorylation. The first two of these pathways are associated intimately with the assembly of microtubules, but relatively little is known about physiological functions for any of these modified tubulins. Most α-tubulin isotypes are synthesized with a Glu-Tyr dipeptide at the C terminus (Tyr-tubulin), but the tyrosine can be removed by a specific tubulin carboxypeptidase after assembly into a microtubule, leaving a terminal glutamate (Glu-tubulin). When microtubules containing detyrosinated α-tubulins are disassembled, the liberated α-tubulins are retyrosinated rapidly by a specific tubulin tyrosine ligase. The result is that microtubules that have been assembled for an extended period of time will tend to be rich in Glu-tubulin. The tyrosination state of α-tubulin does not affect its assembly–disassembly kinetics *in vitro*, but evidence suggests that detyrosination may affect the interaction of microtubules with other cellular structures. In parallel with detyrosination, α-tubulins are also substrates for a specific acetylation reaction. Acetylation of tubulin was initially described for flagellar tubulins, but subsequent work demonstrated that this modification was widespread in neurons and many other cell types. Because the acetylase acts preferentially on α-tubulin assembled into microtubules, long-lived or stable microtubules tend to be rich in acetylated α-tubulin. However, the distribution of microtubules rich in acetylated tubulin may not be identical with that of Glu-tubulin. Acetylated α-tubulin is also deacetylated rapidly upon disassembly of microtubules, although acetylation does not alter the stability of microtubules *in vitro*.

In contrast with these modifications, tubulin phosphorylation involves a β-tubulin and appears to be restricted to an isotype expressed preferentially in neurons and neuron-like cells. A variety of kinases have been shown to phosphorylate tubulin *in vitro*, but the endogenous kinase has not been identified. The effect of phosphorylation on assembly is unknown, but phosphorylation is upregulated during neurite outgrowth. As with α-tubulin modifications, the physiological role of phosphorylation on neuronal β-tubulin has yet to be determined. A variety of additional posttranslational modifications have been reported, but their significance

and distribution in the nervous system are not well documented.

The biochemical diversity of microtubules is increased through the association of different MAPs with different populations of microtubules (Table 4.2). The significance of microtubule diversity is not completely understood, but it may include functional differences as well as variations in assembly and stability. In particular, MAP composition may be used to define specific neuronal domains. For example, one type of MAP, MAP-2, appears to be restricted to dendritic regions of the neuron, whereas another class of MAPs, tau proteins, are modified differentially in axons. A recently identified isoform of MAP-2 is similar to MAP-2c but includes an additional repeat within the microtubule-binding site; hence it is known as 4-repeat MAP-2c, or MAP-2d.57. Oligodendrocyte progenitors transiently express this novel isoform of MAP-2c in their cell bodies but not in their processes, suggesting that MAP-2d might have a role separate from its known capacity to bundle microtubules (Vouyiouklis & Brophy, 1995).

MAPs in nervous tissue fall into two heterogeneous groups: tau proteins and high molecular weight MAPs. Tau proteins have been the subject of intense interest because posttranslationally modified tau proteins are the primary polypeptide constituents of neurofibrillary tangles from the brains of Alzheimer patients. Tau proteins appear to be neuronal MAPs, although reports of tau immunoreactivity outside neurons have appeared. Tau proteins bind to microtubules during assembly–disassembly cycles with a constant stoichiometry and can promote microtubule assembly and stabilization. However, recent evidence suggests a role for tau protein as a scaffold for signaling molecules and localizing these pathways to the microtubule that may be more important than stabilization in the mature neuron. Tau exists in a number of molecular weight isoforms that vary with region of the nervous system and developmental stage. For example, tau proteins in the adult CNS are typically from 60 to 75 kDa, whereas PNS axons contain a higher molecular mass tau of approximately 100 kDa. The different isoforms of tau protein are generated from a single mRNA by alternative splicing, and additional heterogeneity is produced by phosphorylation.

In contrast with tau MAPs, high molecular weight MAPs are a diverse group of largely unrelated proteins found in a variety of tissues, although some are brain specific. Most have apparent molecular mass of 1300 kDa or more and form side arms protruding from the walls of microtubules. Many of them may participate in microtubule assembly and cytoskeletal organization but may also serve as scaffolds. Traditionally, the high molecular weight MAPs comprise five polypeptides: MAPs 1a, 1b, 1c, 2a, and 2b. MAP-2 proteins are closely related and are located primarily in dendrites. In contrast, the three polypeptides known as MAP-1 are unique polypeptides with little sequence homology. MAP-1c is a cytoplasmic form of dynein (see the section on molecular motors later in this chapter). MAPs 1a and 1b are widespread and appear to be regulated developmentally. MAPs 1a, 1b, and 2 are all thought to play important roles in stabilizing and organizing the microtubule cytoskeleton.

TABLE 4.2 Major Microtubule Proteins and Microtubule Motors in Mammalian Brain

	Location and Function
Tubulins	
α-and β-tubulins	Neurons, glia, and nonneuronal cells except mature mammalian erythrocytes. Multigene family with some genes expressed preferentially in brain, whereas others are ubiquitous. Primary structural polypeptides of microtubules
γ-Tubulin	Present near microtubule-organizing center in all microtubule-containing cells. Needed for nucleation of microtubules
	Microtubule-Associated Proteins (MAPs)
MAP-1a/1b	Widely expressed in neurons and glia, including both axons (MAP-1b) and dendrites (MAP-1a); developmentally regulated phosphoproteins.
MAP-2a/2b MAP-2c	Dendrite-specific MAPs. The smaller MAP-2c is regulated developmentally, becoming restricted to spines in adults, whereas 2a and 2b are major phosphoproteins in adult brain
LMW tau HMW tau	Tau proteins are enriched in axons with a distinctive phosphorylation pattern. A single tau gene is alternatively spliced to give multiple isoforms. LMW tau is found in the CNS and HMW tau is seen in the mature PNS
Katanin and Spastin	Microtubule severing proteins that release MTs from the microtubule organizing center (K) or play a role in branching of neurites (K and S)
	Motor Proteins
Kinesin-1	Plus-end directed motors associated with membrane-bound organelles to move them along microtubules in fast axonal transport.
Kinesin-related proteins	A diverse set of motor proteins with a kinesin-related motor domain and varied tails. Many are regulated developmentally and some are mitotic motors, restricted to dividing cells. May be plus- or minus-end-directed motors
Axonemal dynein	A set of minus-end-directed microtubule motors associated with cilia and flagella, such as ependymal cells.
Cytoplasmic dynein (MAP-1c)	Cytoplasmic form of minus-end-directed microtubule-based motor may be involved in the axonal transport of both organelles or cytoskeletal elements

In most cell types, cytoplasmic microtubules appear to be relatively dynamic structures, although stable microtubules or microtubule segments are found in all cells. In nonneuronal cells such as astrocytes and other glial cells, microtubules are typically anchored in centrosomal regions that serve as microtubule organizing centers. As a result, their cytoplasmic microtubules are oriented so that plus ends are distal to the cell center. The biochemistry of microtubule organizing centers is not fully understood, but they contain a novel tubulin subunit, γ-tubulin, which is thought to function as a nucleating site for microtubules. In contrast, dendritic and axonal microtubules of neurons are not continuous with the microtubule organizing center, so alternate mechanisms must exist for stabilization and organization of these microtubules. The situation is complicated further by the fact that dendritic and axonal microtubules differ in both composition and organization. Studies show that both axonal and dendritic microtubules are nucleated at the microtubule-organizing center but are subsequently released for delivery to the appropriate compartment through the action of the microtubule severing protein, katanin (Roll-Mecak & McNally, 2010). Surprisingly, axonal and dendritic compartments are not equivalent. There are two striking differences. First, MAPs of dendritic and axonal microtubules are different in both identity and phosphorylation state. Second, microtubule orientation in axons is similar to that seen in other cell types with the plus end distal, but microtubules in dendrites may exhibit both polarities. Perhaps due to these differences, dendritic microtubules are less likely to be aligned with one another and appear less regular in their spacing.

Stabilization of axonal and dendritic microtubules is essential because of the volume of cytoplasm and distance from sites of protein synthesis for tubulin. Because microtubules play critical roles in both dendritic and axonal function, mechanisms to ensure their proper extent and organization must exist. A common side effect of one class of antineoplastic drugs, the vinca alkaloids, underscores the importance of microtubule stability in axons. Vincristine and other vinca alkaloids act by destabilizing spindle microtubules, but dosage must be monitored carefully to prevent the development of peripheral neuropathies due to loss of axonal microtubules.

Axonal microtubules contain a particularly stable subset of microtubule segments that are resistant to depolymerization by antimitotic drugs, cold, and calcium. The stable microtubule segments are biochemically distinct and may constitute more than half of the axonal tubulin. Stable domains in microtubules may serve to regulate the axonal cytoskeleton by nucleating and organizing microtubules as well as stabilizing them. The biochemical basis of microtubule stability is not completely understood but probably includes posttranslational modification of the tubulin, the presence of stabilizing proteins, or both. There are indications that levels of cold-insoluble tubulin correlate with axonal plasticity. In contrast, relatively little is known about the regulation of dendritic microtubules, but local synthesis of MAP-2 in dendrites may play a role in regulating their stability.

Microfilaments and the Actin-Based Cytoskeleton Are Involved in Intracellular Transport and Cell Movement

The actin cytoskeleton is universally present in eukaryotes, although actin microfilaments are most familiar as the thin filaments of skeletal muscle. Microfilaments (Table 4.3) play a critical role in contractility for both muscle and nonmuscle cells. Actin and its contractile partner myosin are particularly abundant in nervous tissue relative to other nonmuscle tissues (Letourneau, 2009). In fact, one of the earliest descriptions of nonmuscle actin and myosin was in brain (Berl, Puszkin, & Nicklas, 1973). In neurons, actin microfilaments are most abundant in presynaptic terminals, dendritic spines, growth cones, and the subplasmalemmal cortex. Although concentrated in these regions, microfilaments are also present throughout the cytoplasm of both neurons and glia in the form of short filaments from 4 to 6 nm in diameter and from 400 to 800 nm in length.

As with tubulin, multiple actin genes exist in both vertebrates and invertebrates. Four α-actin human genes have been cloned. Each of these α-actin genes is expressed specifically in a different muscle cell type (skeletal, cardiac, vascular smooth, and enteric smooth muscle). In addition to α-actins, two nonmuscle actin genes (β-and γ-actin) are present in humans. β-Actin

TABLE 4.3 Selected Proteins of the Microfilament Cytoskeleton in Brain

Actins
α-actin (smooth and skeletal muscle)
β-actin and γ-actin (neuronal and nonneuronal cells)
Actin monomer-binding proteins
Profilin
Thymosin 4 and 10
Capping proteins
EB1, EB3
Ezrin/radixin/moesin
Schwannomin/merlin
Gelsolin and other microfilament severing proteins
Gelsolin
Villin
Cross-linking and bundling proteins
Spectrin (fodrin)
Dystrophin, utrophin, and related proteins
α-Actinin
Tropomyosin
Myosins I, II, V, VI, VII

and γ-actin genes are expressed ubiquitously, and both are abundant in nervous tissue. The functional significance of these different genetic isotypes is not clear because the actins are highly conserved proteins. Across the range of known actin sequences, the amino acids are identical at approximately two of three positions. Even the positions of introns within different actin genes are highly conserved across many species and genes. Despite this high degree of conservation, differences in the distribution of specific isotypes within a single neuron have been reported. For example, β-actin may be enriched in growth cones. The prominent actin bundles seen in fibroblasts and some other nonneuronal cells in culture are not characteristic of neurons, and most neuronal actin microfilaments are less than 1 μm in length.

Many microfilament-associated proteins have been described in nervous tissue (myosin, tropomyosin, spectrin, α-actinin, etc.), but less is known about their distribution and normal function in neurons and glia. Myosins and myosin-associated proteins are considered later in the section on molecular motors, but several categories of actin-binding proteins can be defined (Table 4.3). Monomer actin-binding proteins such as profilin and thymosin β4 or β10 are abundant in the developing brain and are thought to help regulate the amount of actin assembled into microfilaments by sequestering actin monomers. These monomers can be mobilized rapidly in response to appropriate signals. For example, phosphatidylinositol 4,5-bisphosphate causes the actin–profilin complex to dissociate, freeing the monomer for microfilament assembly. Such regulation may play a key role in growth cone motility, where actin assembly is an important mechanism for filopodial extension.

Several proteins that can cap actin microfilaments, serving to anchor them to other structures or to regulate microfilament length, have been identified. The ezrin–radixin–moesin gene family encodes barbed-end capping proteins that are concentrated at sites where the microfilaments meet the plasma membrane, suggesting a role in anchoring microfilaments or linking them to extracellular components through membrane proteins. A mutation in a member of this family expressed in Schwann cells, merlin or schwannomin, is responsible for the human disease neurofibromatosis type 2. Development of numerous tumors with a Schwann cell lineage in neurofibromatosis type 2 suggests that this microfilament-binding protein acts as a tumor suppressor.

Whereas some membrane proteins can interact directly with actin microfilaments of the membrane cytoskeleton, others interact with the actin cytoskeleton through intermediaries such as spectrin. Proteins such as spectrin (fodrin), α-actinin, and dystrophins cross-link, or bundle, microfilaments, giving rise to higher order complexes. Spectrin is enriched in the cortical membrane cytoskeleton and is thought to have a role in the localization of integral membrane proteins such as ion channels and receptors. Dystrophin is the best-known member of a family of related proteins that all appear to be essential for the clustering of receptors in muscle and nervous tissue. A mutation in dystrophin is responsible for Duchenne muscular dystrophy. Positioning of integral membrane proteins on the cell surface is likely to be an essential function of the actin-rich membrane cytoskeleton, acting in concert with a class of proteins that contain the protein-binding module, the PDZ domain, a motif that typically interacts with the C-terminus of transmembrane proteins.

Members of the gelsolin family have multiple activities. They cannot only cap the barbed end of a microfilament, but also sever microfilaments and nucleate microfilament assembly under some circumstances. These severing-capping proteins may be essential for reorganizing the actin cytoskeleton. Because gelsolin severing activity is Ca^{2+} activated, it may provide a mechanism for altering the membrane cytoskeleton in response to Ca^{2+} transients. Other second messengers, such as phosphatidylinositol 4,5-bisphosphate, may also serve as regulators of gelsolin function, suggesting an interplay between different classes of actin-binding proteins such as gelsolin and profilin. Oligodendrocytes are the only nonneuronal cells in the CNS that express significant amounts of the actin-binding and microfilament-severing protein gelsolin (Tanaka & Sobue, 1994).

Proteins with other functions may also interact directly with actin or actin microfilaments. For example, the enzyme DNase I binds actin tightly, inhibiting both DNase activity and actin assembly. The physiological function of this interaction is unclear, but it has proved a useful tool for probing actin structure and function. Some membrane proteins, such as the epidermal growth factor receptor, bind actin microfilaments directly, which may be important in anchoring these membrane components at a particular location on the cell surface. Other cytoskeletal structures may have specific interactions. Both MAP-2 and tau microtubule-associated proteins have been shown to interact with actin microfilament *in vitro* and have the potential to mediate interactions between microtubules and microfilaments. Finally, the synaptic vesicle-associated phosphoprotein, synapsin I, has a phosphorylation-sensitive interaction with microfilaments that appears to be important for the targeting and storage of synaptic vesicles in the presynaptic terminal. Many of these interactions have been defined by *in vitro*–binding studies, and their physiological significance is not always clearly established. However, there is little doubt that interactions occur between the actin cytoskeleton and a variety of other cellular structures.

The presence of actin as a major component of both pre- and postsynaptic specializations, as well as in the growth cone, gives the actin cytoskeleton special significance in

the nervous system. The enrichment of microfilaments and associated proteins in the membrane cytoskeleton means that they are the cytoskeletal components most subject and most responsive to changes in the local external environment of the neuron.

Microfilaments also play a critical role in positioning the various receptors and ion channels at specific locations on the neuronal surface, such as the initial segment (Grubb & Burrone, 2010). Although many studies have emphasized enrichment of the microfilament cytoskeleton at the plasma membrane, microfilaments are also abundant in the deep cytoplasm. In many respects, the microfilaments may be best regarded as a uniquely plastic component of the neuronal cytoskeleton that plays a critical role in local trafficking of both cytoskeletal and membrane components.

Intermediate Filaments Are Prominent Constituents of Nervous Tissue

Intermediate filaments of the nervous system appear as solid, ropelike fibrils from 8 to 12 nm in diameter that may be many micrometers long (Herrmann, Bar, Kreplak, Strelkov, & Aebi, 2007). Intermediate filament proteins constitute a superfamily of five classes, which have distinctive patterns of expression specific to cell type and developmental stage (Table 4.4).

Type I and type II intermediate filament proteins are keratins, which are hallmarks of epithelial cells. Keratins are not associated with nervous tissue and will not be considered further here. In contrast, all nucleated cells contain type V intermediate filament proteins, the nuclear lamins. Lamins are encoded by the most evolutionarily divergent of the intermediate filament genes, with a distinctive pattern of introns and exons, as well as having a different polypeptide domain structure. Intermediate filaments in the nervous system are all produced by either type III or type IV intermediate filament proteins.

Type III intermediate filaments are a diverse family that includes, among others, vimentin (characteristic of fibroblasts and many embryonic tissues such as embryonic neurons) and glial fibrillary acidic protein (GFAP, a marker for astrocytes and Schwann cells). Type III intermediate filament subunits typically have a molecular mass between 45 and 60 kDa and consist of a conserved rod domain and relatively small gene-specific amino- and carboxy-terminal sequences. As a result, intermediate filaments formed from type III subunits form smooth filaments without side arms. Type III polypeptides can form homopolymers but may also coassemble with other type III intermediate filament subunits.

Another type III intermediate filament protein, peripherin, is unique to neurons and may be coexpressed with neurofilament triplet proteins. Peripherin has a characteristic expression during development and regeneration in specific neuronal populations. It has been shown to coassemble with neurofilament triplet proteins both *in vitro* and *in vivo*, where presumably it can substitute for the low molecular weight neurofilament (NFL). However, whether coassembly is generally the case is not known. Physiological roles for neuron-specific type III intermediate filament polypeptides are uncertain. Unlike type IV intermediate filaments, intermediate filaments made from type III subunits tend to disassemble more readily under physiological conditions. Thus, the presence of type III intermediate filament subunit proteins may produce more dynamic structures, which could be important during development or regeneration.

Although other type III intermediate filament proteins are found in the nervous system, they are generally restricted to glia or to neurons at early stages of differentiation. Vimentin is abundant in a wide variety of cells during early development, including both glioblasts and neuroblasts. Some Schwann cells and astrocytes contain vimentin. Curiously, mature oligodendrocytes do not appear to have any cytoplasmic intermediate filaments; an exception to the general rule that metazoan cells contain all three classes of cytoskeletal structures. Oligodendrocyte precursors do, however, express vimentin and may express GFAP transiently.

In neurons, intermediate filaments typically have side arms that limit packing density, whereas glial intermediate filaments lack side arms and may be very tightly packed. Neuronal intermediate filaments have

TABLE 4.4 Intermediate Filament Proteins of the Nervous System

Class and Name	Cell Type
Types I and II	
Acidic and basic keratins	Epithelial and endothelial cells
Type III	
Glial fibrillary acidic protein	Astrocytes and nonmyelinating Schwann cells
Vimentin	Neuroblasts, glioblasts, fibroblasts, etc.
Desmin	Smooth muscle
Peripherin	Asubset of peripheral and central neurons
Type IV	
NF triplet (NFH, NFM, NFL)	Most neurons, expressed at highest level in large myelinated fibers
α-Internexin	Developing neurons, parallel fibers of cerebellum
Nestin	Early neuroectodermal cells. The most divergent member of this class; some have classified it as a sixth type
Type V	
Nuclear lamins	Nuclear membranes

an unusual degree of metabolic stability, which makes them well suited to the role of stabilizing and maintaining neuronal morphology. The existence of neurofilaments was established for many years before much was known about their biochemistry or function. Neurofilaments could be seen in early electron micrographs, and many traditional histological procedures visualize neurons as a result of a specific interaction of metals with neurofilaments. Recent work on the biochemistry and molecular genetics of intermediate filaments has illuminated many aspects of their function in the nervous system.

The primary type of intermediate filament in neurons is formed from three subunits, the neurofilament triplet, each encoded by a separate gene. Neurofilament triplet proteins are type IV intermediate filament proteins, which are generally expressed only in neurons and have a characteristic domain structure seen in both primary sequence and gene structure. The polypeptides were initially identified from axonal transport studies. Apparent molecular mass for the neurofilament subunits vary widely across species, but mammalian forms typically range from 180 to 200 kDa for the high molecular weight subunit (NFH), from 130 to 170 kDa for the medium subunit (NFM), and from 60 to 70 kDa for the NFL. Interestingly, Schwann cells in damaged peripheral nerves also transiently express NFM and NFL. Neurofilament subunits are phosphorylated in axons, with NFM and NFH having unusually high levels of phosphorylation. In some species, NFH has 50 or more repeats of a consensus phosphorylation site at its carboxy terminus, and levels of NFH phosphorylation indicate that most of these sites are phosphorylated *in vivo*. This high level of phosphorylation in neurofilament subunit tail domains is a distinctive characteristic of neurofilaments.

A second motif characteristic of neurofilaments is the presence of a glutamate-rich region in the tail adjacent to the core rod domain. This glutamate region has particular significance for neuroscientists because it appears to be the basis for the reaction of the classic neurofibrillary silver stains for neurons. These stains were first introduced in the late nineteenth century and have been used extensively by neurohistologists and neuroanatomists from Ramon y Cajal's time to the present day. However, the molecular basis of these neurofibrillary stains was not known until 1968, when F. O. Schmitt showed that neurofibrils were formed by 10-nm-diameter neurofilaments. Remarkably, the ability of isolated neurofilament subunits to react with silver histological stains is retained even after separation in gel electrophoresis for neurofilaments from organisms as diverse as humans, squid, and the marine fanworm, *Myxicola*. Conservation of the glutamate-rich domain suggests both an important functional role for this motif and the early divergence of neurofilaments from the other intermediate filament families.

Neurofilaments formed from neurofilament triplet proteins play a critical role in determining axonal caliber. As mentioned earlier, neurofilaments have characteristic side arms, unique among intermediate filaments; these side arms are formed by NFM and NFH carboxy-terminal regions. Although all three neurofilament subunits contribute to the neurofilament central core, the side arms are formed only by the NFM and NFH subunits. Phosphorylation of NFH and NFM side arms alters charge density on the neurofilament surface, repelling adjacent neurofilaments with a similar charge. The high density of surface charge due to phosphate groups on neurofilaments makes it difficult to imagine a stable interaction between neurofilaments and other structures of like charge. Although many reports refer to cross bridges between neurofilaments, direct studies of interactions between neurofilaments provide little evidence of stable crosslinks between neurofilaments or between neurofilaments and other cytoskeletal structures. However, dynamic interactions between neurofilaments and cellular structures or proteins may be critical for many aspects of neurofilament function and metabolism.

Alteration in expression levels for neurofilament subunits or mutations in neurofilament genes can lead to specific neuropathologies. Overexpression of genes encoding normal NFH or expression of some mutant NFL genes in transgenic mouse models leads to the accumulation of neurofilaments in the cell body and proximal axon of spinal motor neurons. These accumulations are similar to those seen in amyotrophic lateral sclerosis and related motor neuron diseases, leading to the hypothesis that the disruption of normal neurofilament function is a common intermediate in the pathogenesis of motor neuron disease. Similarly, an early indicator of neuropathies caused by neurotoxins such as acrylamide and hexanedione is the accumulation of neurofilaments in either proximal or distal regions of the axon. Disruption of neurofilament organization is a hallmark of pathology for many degenerative diseases of the nervous system, particularly those affecting large myelinated axons such as those of spinal motor neurons. Although altering neurofilament organization can produce pathology, the question of whether neurofilament defects are a primary event in pathogenesis or a manifestation of an underlying metabolic pathology remains controversial in most cases.

Another member of the type IV intermediate filament family, α-internexin, is also expressed only in neurons. Unlike the triplet proteins, α-internexin is expressed preferentially early in development of the nervous system and then disappears from most neurons during maturation. Intermediate filaments containing α-internexin do

persist in portions of the adult nervous system, such as the branched axons of granule cells in the cerebellar cortex. Evidence show that α-internexin can coassemble with members of the neurofilament triplet, but it also forms homopolymeric filaments. The primary sequence of α-internexin has features in common with both NFL and NFM that are thought to form the basis for assembly properties distinct from those of other type IV intermediate filaments.

The final type of intermediate protein present in the nervous system is nestin, which is expressed transiently during early development. Nestin is also expressed in Schwann cells and in the progenitors of oligodendrocytes, which appear late in the development of the embryonic nervous system. Remarkably, nestin appears to be expressed almost exclusively in ectodermal cells after commitment to the neuroglial lineage, but prior to terminal differentiation. At 1250 kDa, nestin is the largest intermediate filament subunit and is the most divergent in sequence with several distinctive features, leading some to classify nestin as a sixth type of intermediate filament protein, whereas others group it with type IV intermediate filaments. Relatively little is known about the assembly properties of nestin *in vivo* or the physiological function of nestin intermediate filaments in neuroectodermal cells.

How Do the Various Cytoskeletal Systems Interact?

Each class of cytoskeletal structures may be found without the others in some cellular domains, but all three classes—microtubules, microfilaments, and intermediate filaments—coexist in many domains. As a result, they inevitably interact. This is not to say that individual cytoskeletal elements are necessarily cross-linked to one another. As mentioned earlier, microtubules and neurofilaments have highly phosphorylated side arms that project from their surfaces. The high density of negative charge on the surface tends to repel structures with a like charge such as other microtubules and neurofilaments. This does not mean that microtubules and neurofilaments do not interact with each other, but it does suggest that such interactions may be transient.

One location containing longer microfilaments and more elaborate organization is the growth cone, which contains bundles of microfilaments in the filopodia, as well as a more dispersed actin network. Neurofilaments are largely excluded from the growth cone, typically extending no further than the neck of the growth cone. In contrast, microtubules and microfilaments play complementary roles in the growth cone itself. Microfilaments are critical in sprouting but appear less critical for elongation, at least over short distances. Disruption of microtubules in the distal neurite does not affect sprouting but does inhibit neurite elongation. It seems likely that microtubules and microfilaments also act synergistically in the function of the axon initial segment. The tips of microtubules are often associated with proteins that appear to regulate their interaction with sub-plasma membrane structures. At the axon initial segment the proteins EB1 and EB3 appear to link microtubules with ankyrin-G, a key interactor with the spectrin-actin cytoskeleton and that is known to play a crucial role in the assembly of this critical neuronal domain.

Summary

The intracellular framework that gives shape to the neuron and glial cell is the cytoskeleton, a complicated set of filaments and tubules and their associated proteins. These organelles are responsible as well for intracellular movement of materials and, during development, for cell migration and plasma membrane extension within nervous tissue.

MOLECULAR MOTORS IN THE NERVOUS SYSTEM

Until 1985, our knowledge of molecular motors in vertebrate cells of any type was restricted to myosins and flagellar dyneins. Myosins had been identified in nervous tissue, but their functions were uncertain. Because the preponderance of evidence indicated that fast axonal transport was microtubule based, there was considerable interest in dyneins in cell cytoplasm. Despite a number of studies, no evidence for a functional cytoplasmic dynein emerged. Worse yet, the characteristic properties of fast organelle movements appeared inconsistent with both myosins and dyneins. Over the past decade, however, we have developed a good but still incomplete understanding of how these motors may work inside cells.

Myosins and dyneins can be distinguished pharmacologically by their differential susceptibility to inhibitors of ATPase activity, but the spectrum of inhibitors active against fast axonal transport fails to match the properties of either myosin or dynein. The most striking difference between inhibitor effects on axonal transport and on myosin or dynein motors was seen in the effect of a non-hydrolyzable analog of ATP. Adenylyl-imidodiphosphate (AMP-PNP) is a weak competitive inhibitor of both myosin and dynein, requiring a 10- to 100-fold excess of analog. In contrast, within minutes of AMP-PNP perfusion into isolated axoplasm, both anterograde and retrograde axonal transport stop. Inhibition by AMP-PNP occurs even in the presence of stoichiometric concentrations of

ATP. Organelles moving in both directions freeze in place and remain attached to microtubules. AMP-PNP weakens the interaction of myosin with microfilaments and of dynein with microtubules, but stabilizes the binding of membrane-bound organelles to microtubules. Thus, the effects of AMP-PNP indicate that movement of membrane-bound organelles in fast axonal transport must require another type of motor, distinct from myosins and dyneins.

The effects of AMP-PNP both demonstrated the existence of a new type of mechanochemical ATPase and provided a basis for identifying its constituent polypeptides. Binding of ATPase to microtubules should be increased by AMP-PNP and decreased by ATP. Polypeptides meeting this criterion were soon identified. The new ATPase was named kinesin, based initially on an ability to move microtubules across glass coverslips as first described in axoplasmic extracts. Studies soon established that kinesin was a microtubule-activated ATPase with minimal basal activity. This combination of ATPase activity and motility *in vitro* confirmed that kinesin was a new class of microtubule-based motor (Brady & Sperry, 1995).

Kinesin has now been identified in a variety of organisms and tissues, leading to an extensive characterization of many biochemical, pharmacological, immunochemical, and molecular properties. Electron microscopic and biophysical analyses reveal kinesin as a long, rod-shaped protein, approximately 80 nm in length. Neuronal kinesin is a heterotetramer with two heavy chains (molecular mass 115–130 kDa) and two light chains (62–70 kDa). Localization of antibodies specific for kinesin subunits by high-resolution electron microscopy of bovine brain kinesin indicates that the two heavy chains are arranged in parallel, forming the heads and much of the shaft, whereas light chains are localized to the fan-shaped tail region (Fig. 4.5).

A variety of approaches have demonstrated that the ATP-binding and microtubule-binding domains of kinesin are in the head regions of the heavy chains, whereas the light chains in the tail region of kinesin appear to bind to membranes. When *in vitro* motility assays are employed for analysis of brain kinesins, movements are directed toward the plus ends of microtubules. Because axonal microtubules are oriented with their plus ends distal from the cell body, this movement would be appropriate for a motor that moves organelles in the anterograde direction.

Neuronal kinesin appears associated with a variety of membrane-bound organelles, including synaptic vesicles, mitochondria, coated vesicles, and lysosomes. The interaction of kinesin and other molecular motors with membrane surfaces is not well understood. In the case of kinesin, the interaction is thought to involve the light chains of kinesin along with the carboxy termini of the heavy chains.

Kinesins have now been shown to be a family of related proteins with a highly conserved domain that includes ATP- and microtubule-binding domains (Hirokawa, Noda, Tanaka, & Niwa, 2009). Many of these kinesin-related polypeptides appear associated with cell division, and kinesin-related proteins in vertebrate tissues are not well characterized. However, multiple members of the kinesin superfamily are expressed in both adult and developing brains. This proliferation of motor proteins has dramatically altered the questions being asked about motor function in the brain. The discovery that ncd, a kinesin-related protein from *Drosophila*, can move structures toward the minus end of microtubules increases the number of potential functions that kinesin family members might serve in nervous tissue still further, perhaps including a role in retrograde transport. Further study is needed to establish specific functions for each member of the kinesin superfamily expressed in neurons or glial cells.

As an indirect result of the discovery of kinesin, one of the high molecular weight microtubule-associated proteins of brain, MAP-1c, was found to be the long sought cytoplasmic form of dynein (Allan, 2011). Both MAP-1c dynein and kinesin can be isolated from bovine brain by incubation of microtubules with nucleotide-free soluble extracts. Both are bound to microtubules under these conditions and released by ATP. MAP-1c dynein moved microtubules *in vitro* with a polarity opposite that seen with kinesin and was identified as a two-headed cytoplasmic dynein using both structural and biochemical criteria. Concurrently, a similar protein was identified in nematodes.

MAP-1c dyneins form a 40-nm-long complex of molecular mass 1.6×10^6 Da, which includes two heavy chains and a number of light chains (Figs. 4.8 and 4.9). Less information is available about the distribution and properties of MAP-1c dyneins than about kinesin. Immunocytochemical studies in nonneuronal cells showed immunoreactivity on mitotic spindles. In addition, a punctate pattern of immunoreactivity also present in interphase cells was thought to be due to dynein bound to membrane-bound organelles. Dyneins are widely thought to be the motor for fast retrograde axonal transport but are also a candidate for a motor in slow axonal transport.

Myosins from muscle were the first molecular motors identified, but in recent years interest in nonmuscle myosins has increased (Foth, Goedecke, & Soldati, 2006). Nonmuscle myosins may be categorized as belonging to one of eight classes, but only a subset has been clearly demonstrated in the nervous system. Nonmuscle myosins in this subset share considerable homology in their motor domains but diverge widely in other domains.

The most familiar of the myosins are the myosin II proteins (Fig. 4.6), which are found in the thick filaments

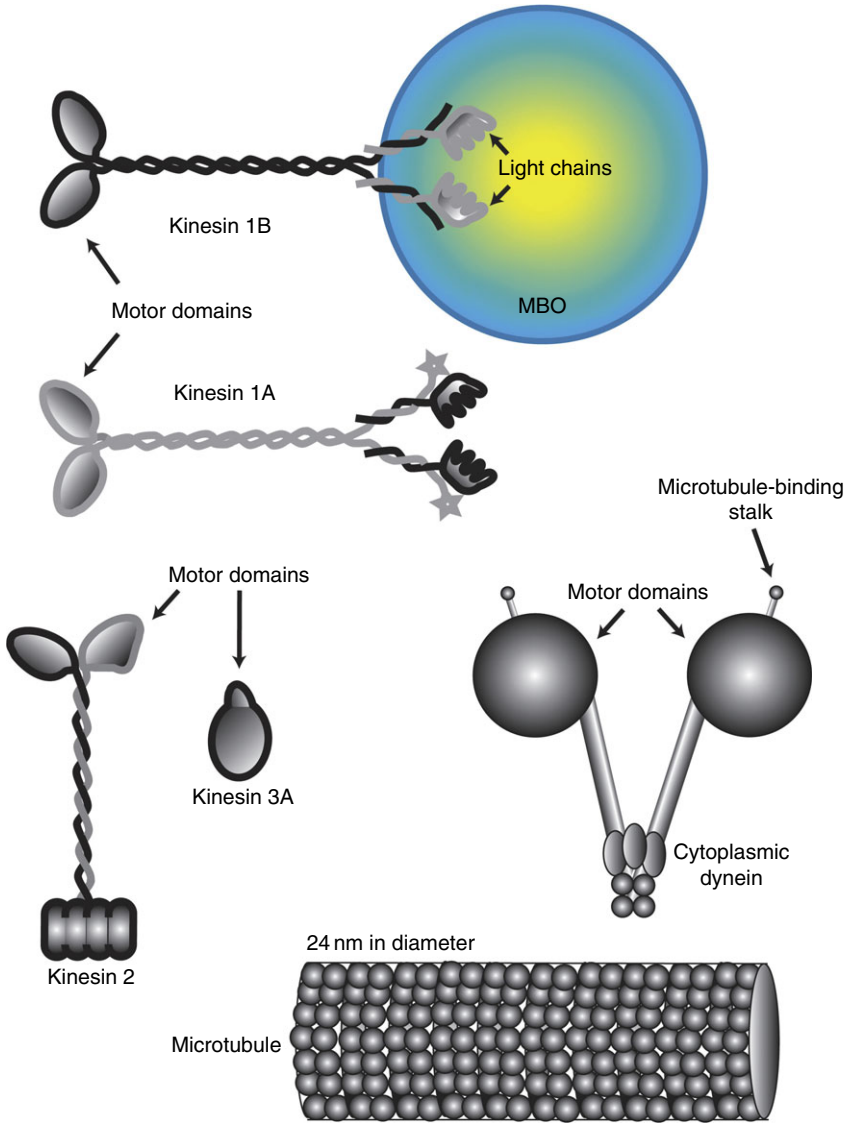


FIGURE 4.5 Examples of microtubule motor proteins in the mammalian nervous system. The first microtubule motor identified in nervous tissue was a kinesin 1, but showed 3 kinesin 1 genes including a neuron-specific form (kinesin 1A). Motor domains are well conserved by tail domains and appear to be specialized for interaction with various targets, such as different membrane-bound organelles. After the sequence of the kinesin heavy chain was established, the presence of additional genes that contained sequences homologous to the motor domain of kinesin was soon recognized. The molecular organization of these various motor proteins is diverse, including monomers (Kinesin 3A), trimers (Kinesin 2), and tetramers (ubiquitous and neuron-specific kinesins). Cytoplasmic dynein may interact with membrane-bound organelles and cytoskeletal structures. Genetic methods have established that there may be more than 40 kinesin-related genes and 16 dynein heavy chains genes in a single organism.

of smooth and skeletal muscle but are also present in nonmuscle cells. Two heavy chains of myosin II form a dimer that may interact with other myosin II dimers to form bipolar filaments. Under tissue culture conditions, many cells contain bundles of actin microfilaments, known as stress fibers, that exhibit a characteristic distribution of myosin II into distinct patterns that may be sarcomeric equivalents, but stress fibers are not apparent in neurons and other cells of the nervous system *in situ*. However, bipolar thick filaments assembled from myosin II dimers can be isolated from nervous tissues. Many of the cellular contractile events described in nonneuronal cells, such as the contractile ring in mitosis, are thought to include myosin II. Although brain myosin II was one of the first nonmuscle myosins to be described, relatively little is known about the function of myosin II in neurons. Myosin II has been localized in the neurites of neurons in primary culture.

Myosin I proteins have a single, smaller heavy chain that does not form filaments but possesses a homologous actinactivated ATPase domain. One exciting aspect of myosin I is its ability to interact directly with membrane

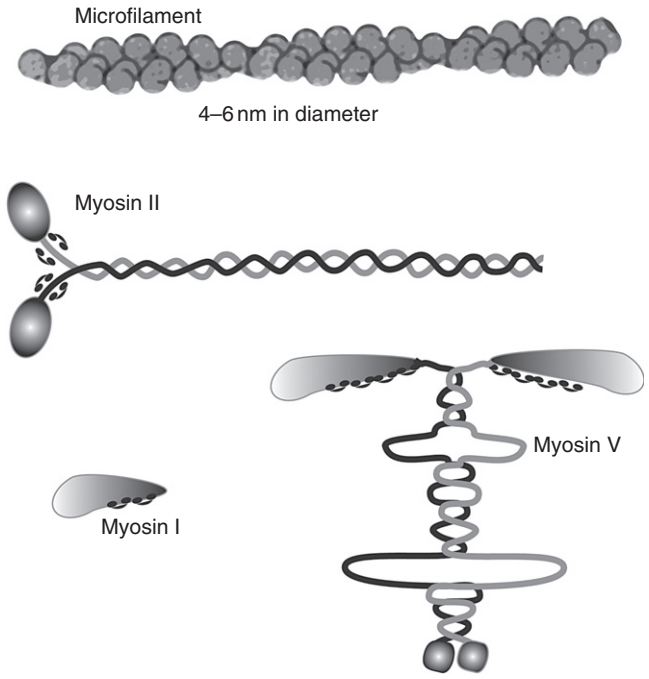


FIGURE 4.6 Examples of myosin motor proteins found in mammalian brain. Myosin heavy chains contain the motor domain, whereas light chains regulate motor function. Myosin II was the first molecular motor characterized biochemically from skeletal muscle and brain. Genetic approaches have now defined more than 15 classes of myosin, many of which are found in brain. Myosin II is a classic two-headed myosin forming thick filaments in nonmuscle cells. Myosin I motors have single motor domains, but may interact with actin microfilaments or membranes. Myosin V has multiple binding sites for calmodulin that act as light chains. Mutations in other classes of myosin have been linked to deafness. Myosins I, II, and V have been detected in growth cones as well as in mature neurons.

surfaces, which may generate movements of plasma membrane components or intracellular organelles. Myosin I has been purified from neural and neuroendocrine tissues. At least three genes in this family have been found in mammals, and multiple forms are present in brain. An interesting aspect of myosin IB is its expression in the stereocilia of hair cells in the cochlea and vestibular system, where it may play a role in mechanotransduction. Both myosin I and myosin II molecules have been proposed to have a role in the motility of lamelipodia at the leading edge of growth cones, but they are also expressed at substantial levels in adult nervous tissue in which growth cones are rare.

The mouse mutation dilute, which affects coat color, was shown to result from a mutation in a gene that encodes a novel myosin heavy chain distinct from both myosins I and II. Similar myosin molecules have been identified in other cell types and organisms and are classified as myosin V. The change in coat color seen in the dilute mouse is due to an inability of skin dendritic pigment cells to deliver the pigment to developing hairs. There are complex neurological deficits in dilute mutants, including seizures that eventually lead to death in severely altered alleles of the dilute mutation. Myosin V was also discovered independently in extracts from chicken and mammalian brain. Myosin V is thought to play a role in the delivery of membrane proteins from the microtubule-based transport system to the plasma membrane. However, neurons in dilute mice without one allele of myosin V clearly develop axons and make connections. The seizures do not begin until early adulthood.

Representatives from two more classes of myosin have been identified in nervous tissue. Genes for a myosin VI and a myosin VIIA have been identified in brain as well as in other tissues. Both have been implicated in forms of congenital deafness. Myosin VI appears to be the gene responsible for Snell's Waltzer deafness, and myosin VIIA has been identified as the gene responsible for a human disease involving both deafness and blindness, Usher syndrome type 1B. Both of these myosins are expressed in mechanosensory hair cells of the cochlea and vestibular apparatus, where they exhibit a different localization from each other and from myosin IB.

The diversity of brain myosins and their distinctive localization suggests that the various myosins may have narrowly defined functions. However, relatively little is known about specific neuronal functions for myosins despite intensive study of myosins in the nervous system. The axonal transport of myosin II-like proteins has been described, but little further progress has been made on the functions of myosin II in the mature nervous system. Even less is known about myosin I in the nervous system. However, myosins likely play roles in growth cone motility, synaptic plasticity, and even neurotransmitter release. There are few instances in our knowledge of neuronal function in which we fully understand the role played by specific molecular motors, but members of all three classes are abundant in nervous tissue. This proliferation of different motor molecules and their isoforms suggests that some physiological activities may require multiple classes of motor molecules.

Summary

The concept is now firmly in place that neurons and glial cells, like other cells, contain molecular motors responsible for moving discrete populations of molecules, particles, and organelles through intracellular compartments.

BUILDING AND MAINTAINING CELLS OF THE NERVOUS SYSTEM

The functional architecture of neurons comprises many specializations in cytoskeletal and membranous components. Each of these specializations is dynamic,

constantly changing, and being renewed at a rate determined by the local environment and cellular metabolism. The processes of axonal transport represent a key to understanding neuronal dynamics and provide a basis for exploring neuronal development, regeneration, and neuropathology. Recent advances are important sources of insight into the molecular mechanisms underlying axonal transport, although many questions remain.

Slow Axonal Transport Moves Soluble Components and Cytoskeletal Structures

Slow axonal transport has two major components, both representing the movement of cytoplasmic constituents (Fig. 4.7). The cytoplasmic and cytoskeletal elements of the axon in axonal transport move at rates at least two orders of magnitude more slowly than fast transport. Slow component a is composed largely of

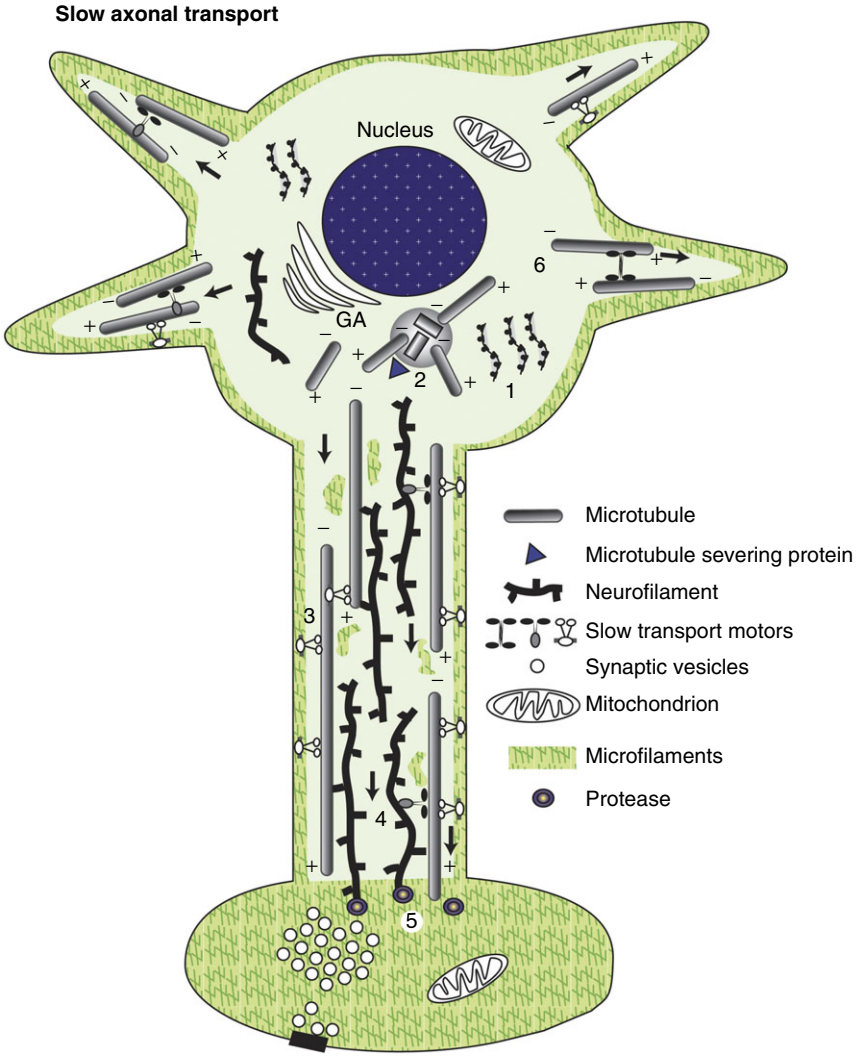


FIGURE 4.7 Slow axonal transport represents the delivery of cytoskeletal and cytoplasmic constituents to the periphery. Cytoplasmic proteins are synthesized on free polysomes and organized for transport as cytoskeletal elements or macromolecular complexes (1). Microtubules are formed by nucleation at the microtubule-organizing center near the centriolar complex (2) and then released for migration into the axon or dendrites. The molecular mechanisms are not as well understood as those for fast axonal transport, but slow transport appears to be unidirectional with no retrograde component. Studies suggest that motors like cytoplasmic dynein may interact with the axonal membrane cytoskeleton to move the microtubules with their plus ends leading (3). Neurofilaments may not be able to move on their own but may hitchhike on microtubules (4). Other cytoplasmic proteins may do the same or may be moved by other motors. Once cytoplasmic structures reach their destinations, they are degraded by local proteases (5) at a rate that allows either growth (in the case of growth cones) or maintenance of steady-state levels. The different composition and organization of the cytoplasmic elements in dendrites suggest that different pathways may be involved in the delivery of cytoskeletal and cytoplasmic materials to the dendrite (6). In addition, some mRNAs are transported into the dendrites, but not into axons.

cytoskeletal proteins, neurofilaments, and microtubule protein. Slow component b is a complex and heterogeneous rate component, including hundreds of distinct polypeptides ranging from cytoskeletal proteins such as actin (and tubulin in some nerves) to soluble enzymes of intermediary metabolism (such as glycolytic enzymes). Many characteristics of axonal transport have been described, and these characteristics provide the foundation for our understanding of mechanisms.

Neurofilaments and microtubules move as discrete cytological structures (Baas & Buster, 2004). Studies on the transport of neurofilament protein indicate that little degradation or metabolism occurs until neurofilaments reach nerve terminals, where they are degraded rapidly. Comparable results have been obtained in studies labeling microtubule protein by radioactivity or fluorescence. Under favorable conditions, movement of individual microtubules can be detected in neurites or growth cones. Both radiolabeling studies and direct observations of individual microtubules indicate that all microtubules and neurofilaments move down the axon, but the motor protein involved is uncertain. Differential metabolism appears to be a key to the targeting of cytoplasmic and cytoskeletal proteins. Proteins with slow degradative rates accumulate and reach higher steady-state concentrations. Alteration of degradation rates changes the steady-state concentration of a protein. The concentration of actin in presynaptic terminals is explained by a slower turnover in terminals relative to neurofilament proteins and tubulin, and inhibition of calpain causes neurofilament rings to appear in presynaptic terminals. Differential turnover may be accomplished by specific proteases or posttranslational modifications that affect the susceptibility to degradation.

The coherent movement of neurofilaments and microtubule proteins provides strong evidence for the "structural hypothesis." For example, pulse-labeling experiments show that radiolabeled neurofilament proteins move as a bell-shaped wave with little or no trailing of neurofilament protein. This fits with the observed stability of neurofilaments under physiological conditions. Thus, any soluble pool of neurofilament subunits is negligible. Similarly, the coherent transport of tubulin and MAPs makes sense only if microtubules are moved because MAPs do not interact with unpolymerized tubulin.

A striking demonstration of microtubule movement can be seen with fluorescent analogs of tubulin (Brown, 2003). When local segments of an axon are photoactivated, short fluorescent microtubule segments can be seen to move down the growing neurite. These microtubules move rapidly (1–2 μm/sec) with frequent pauses that produce a net anterograde transport of microtubules at 1–2 mm/day and similar rates are observed for neurofilaments. When studies of axonal transport using radiolabeled amino acids are combined with direct observations of individual microtubules or neurofilaments by video microscopy, there is little doubt that microtubules and neurofilaments can and do move in the axon as intact, individual cytoskeletal elements.

More recently, similar methods have been used to document the transport of classically soluble, cytoplasmic proteins like glycolytic enzymes and synuclein (Roy, Winton, Black, Trojanowski, & Lee, 2007). As with microtubules and neurofilaments, these movements are intermittent and rapid. The extended pauses lead to a net rate of anterograde transport of approximately 2–4 mm/day, consistent with rates obtained in pulse-chase radiolabel studies.

Fast Axonal Transport Is the Means by Which Membrane Vesicles and their Contents Are Rapidly Moved Long Distances within a Neuron

Early biochemical and morphological studies established that the material moving in fast axonal transport was associated with membrane-bound organelles (Fig. 4.8) (Brady, 1993). A variety of materials could be shown to move in fast transport. In anterograde transport, materials being moved include membrane-associated enzyme activities, neurotransmitters, and neuropeptides. Many of the materials moving down the axon in anterograde transport are returned in retrograde transport, in some cases after modification in the terminal. In addition, a number of exogenous materials taken up in the distal regions of the axon are moved back to the cell body by retrograde transport. Exogenous materials in retrograde transport include neurotrophic factors, such as nerve growth factor, and viral particles invading the nervous system. The uptake of neurotrophic factors may play a critical role in the process of regeneration.

Electron microscopic analysis of materials accumulated at a ligation or crush demonstrated that organelles moving in the anterograde direction were morphologically distinct from those moving in the retrograde direction (Tsukita & Ishikawa, 1980). Consistent with the ultrastructural differences, radiolabel and immunocytochemical studies indicate that there are both quantitative and qualitative differences between anterograde and retrograde moving material. Differences between anterograde and retrograde transport indicate that some processing or repackaging events must occur as part of turnaround in axonal transport. Turnaround processing appears to require a proteolytic event because certain protease inhibitors inhibit turnaround without affecting anterograde and retrograde transport.

Biochemical and morphological approaches resulted in considerable progress toward a description of the

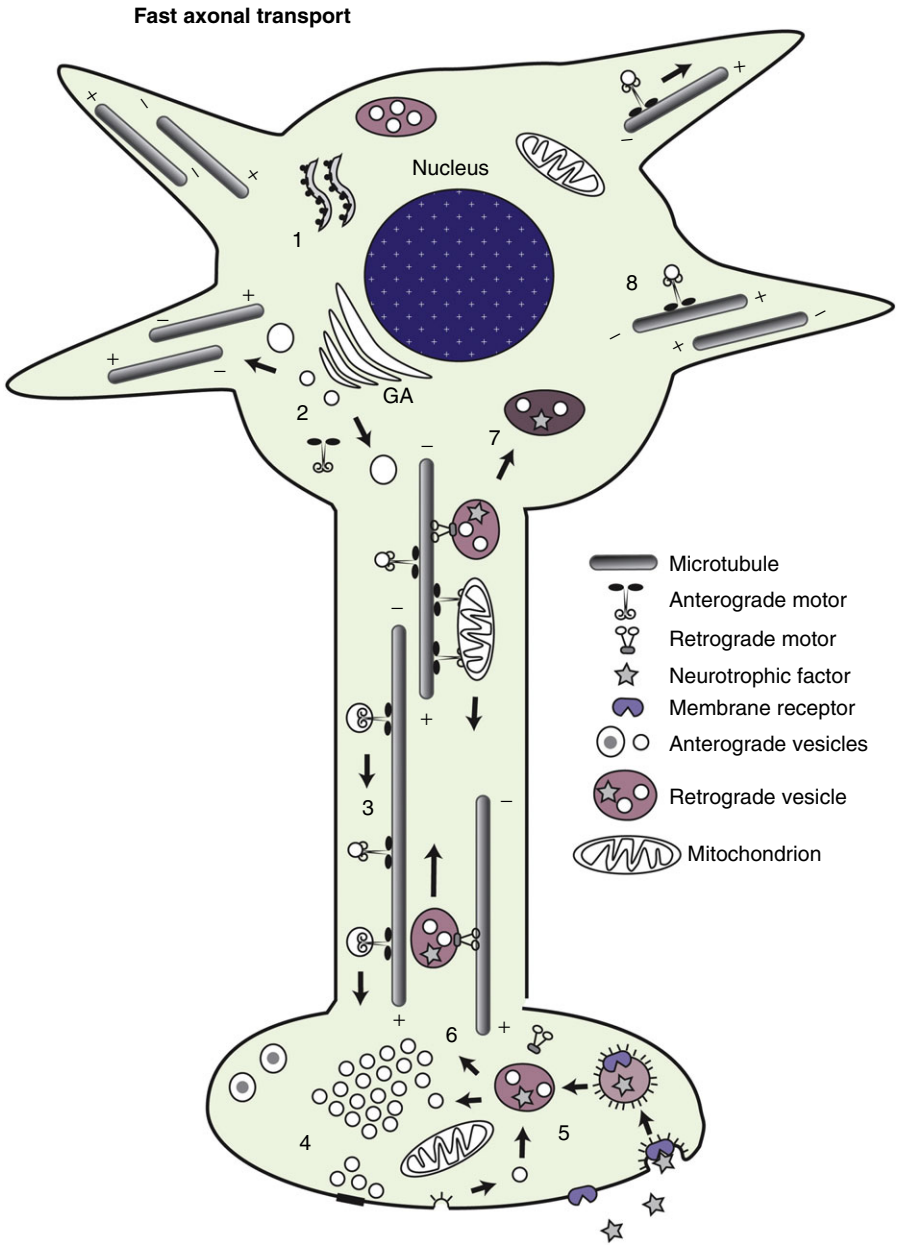


FIGURE 4.8 Fast axonal transport represents the transport of membrane-associated materials, having both anterograde and retrograde components. For anterograde transport, most polypeptides are synthesized on membrane-bound polysomes, also known as rough endoplasmic reticulum (1), and then transferred to the Golgi apparatus for processing and packaging into specific classes of membrane-bound organelles (2). Proteins following this pathway include both integral membrane proteins and secretory polypeptides in the lumen of vesicles. Cytoplasmic peripheral membrane proteins such as kinesins are synthesized on the cytoplasmic or free polysomes. Once vesicles have been assembled and the appropriate motors associate with them, they are moved down the axon at a rate of 100–400 mm per day (3). Different membrane structures are delivered to different compartments and may be regulated independently. For example, dense core vesicles and synaptic vesicles are both targeted for the presynaptic terminal (4), but the release of vesicle contents involves distinct pathways. After vesicles merge with the plasma membrane, their protein constituents are taken up by coated pits and vesicles via the receptor-mediated endocytic pathway and delivered to a sorting compartment (5). After proper sorting into appropriate compartments, membrane proteins are either committed to retrograde axonal transport or recycled (6). Retrograde moving organelles are morphologically and biochemically distinct from anterograde vesicles. These larger vesicles have an average velocity about half that of anterograde transport. The retrograde pathway is an important mechanism for the delivery of neurotrophic factors to the cell body. Material delivered by retrograde transport typically fuses with cell body compartments to form mature lysosomes (7), where most constituents are recycled. However, neurotrophic factors and neurotrophic viruses can act at the level of the cell body. Although evidence shows that vesicle transport also occurs into dendrites (8), less is known about this pocess. Dendritic vesicle transport is complicated by the fact that dendritic microtubules may have mixed polarity.

materials being transported in fast axonal transport but were not suitable for identifying the molecular motors used in translocation. A different technology that permitted direct observation of organelle movements and precise control of experimental conditions was required. Such experiments became possible with the advent of video microscopic techniques (Brady, 1991).

An early use of video-enhanced contrast (VEC) microscopy was to characterize the bidirectional movement of membrane-bound organelles in giant axons from the squid *Loligo pealeii*. Years before, studies had shown that axoplasm could be extruded from the giant axon as an intact cylinder. Properties of the isolated axoplasm had been characterized in some detail, making VEC microscopic analysis of axoplasm a natural choice. Remarkably, fast axonal transport continued unabated in isolated axoplasm for hours. Isolated axoplasm from the giant axon has no plasma membrane or other permeability barriers but can be readily maintained in an active state. Combining VEC microscopy with isolated axoplasm permitted rigorous dissection of the mechanisms for fast axonal transport with the use of biochemical and pharmacological approaches. A number of insights into axonal transport mechanisms have resulted from these studies.

How Is Axonal Transport Regulated?

The diversity of polypeptides in each axonal transport rate component and the coherent movement of proteins having many different molecular weights produce a conundrum. How can so many different polypeptides move down the axon as a group? In theory, one could propose that each protein has a motor of its own. In that case, each rate component might represent movements due to a specific class of motors or to variable affinities for a smaller number of motors. However, the relatively small numbers of motor molecules and the logistical difficulties associated with such a model effectively preclude this possibility.

The structural hypothesis mentioned earlier was formulated in response to the observation that rate components of axonal transport move as discrete waves, each with a characteristic rate and a distinctive composition (Figs. 4.7 and 4.8). The hypothesis is deceptively simple: Axonal transport represents the movement of discrete cytological structures (Brady, 1993). Proteins in axonal transport do not move as individual polypeptides. Instead, they move as part of a cytological structure or in association with a cytological structure. The only assumption made is that a limited number of elements can interact directly with transport motors, so transported material must be packaged appropriately to be moved. The different rate components result from packaging of transported material into different cytologically identifiable structures. In other words, membrane-associated proteins move as membrane-bound organelles (vesicles, mitochondria, etc.), whereas tubulin and MAPs move as microtubules, and neurofilaments move as neurofilaments.

The structural hypothesis does not require movement of a cross-linked microtubule–neurofilament complex in the form of a solid axoplasmic column. Evidence for existence of a cross-linked complex of microtubules and neurofilaments is not compelling in any case. Instead, the hypothesis specifically predicts that individual microtubules and neurofilaments move rather than tubulin dimers or neurofilament monomers. No assumptions are made about higher-order interactions between cytoskeletal structures or membranous structures. Indeed, one variant of the structural hypothesis proposes that cytoskeletal proteins move in the form of small oligomers rather than polymers, although there is no evidence for the presence of such oligomers *in vivo*. For example, to the limits of detection, neurofilaments exist only as the polymer under *in vivo* conditions. Similarly, tubulin interchanges between dimer and polymer with no known oligomeric intermediate. The movement of microtubules, neurofilaments, and complexes of soluble proteins is in fact what is observed (Baas & Buster, 2004; Roy et al., 2007).

Because synthesis of proteins takes place at some distance from many functional domains of a neuron, transport to distal regions of the neuron is necessary, but not sufficient, for proper function. Specific materials must also be delivered to their proper site of utilization and should not be left in inappropriate locations. For example, because a synaptic vesicle has no known function in axons or the cell body, it must be delivered to a presynaptic terminal along with other components necessary for regulated neurotransmitter release. The traditional picture places the presynaptic terminal at the end of an axonal process. Such images imply that a synaptic vesicle need only move along the axonal microtubules until it reaches microtubule ends in the presynaptic terminal. However, many CNS synapses are not at the end of an axon. Numerous terminals are located sequentially along a single axon, making en passant contacts with multiple target cells along the way. Targeting of synaptic vesicles then becomes a more complex problem and targeting ion channels or neurotransmitter receptors to nodes of Ranvier or other appropriate sites on the neuronal surface is equally challenging.

Although the specific details of this targeting are not well understood, a simple model for the targeting of synaptic vesicles serves to illustrate how such targeting may occur (Fig. 4.9). Synapsin I is a cytoplasmic protein enriched in presynaptic terminals that can bind reversibly to both actin microfilaments and synaptic vesicles. Both binding activities are regulated by the phosphorylation of synapsin, and this phosphorylation is increased

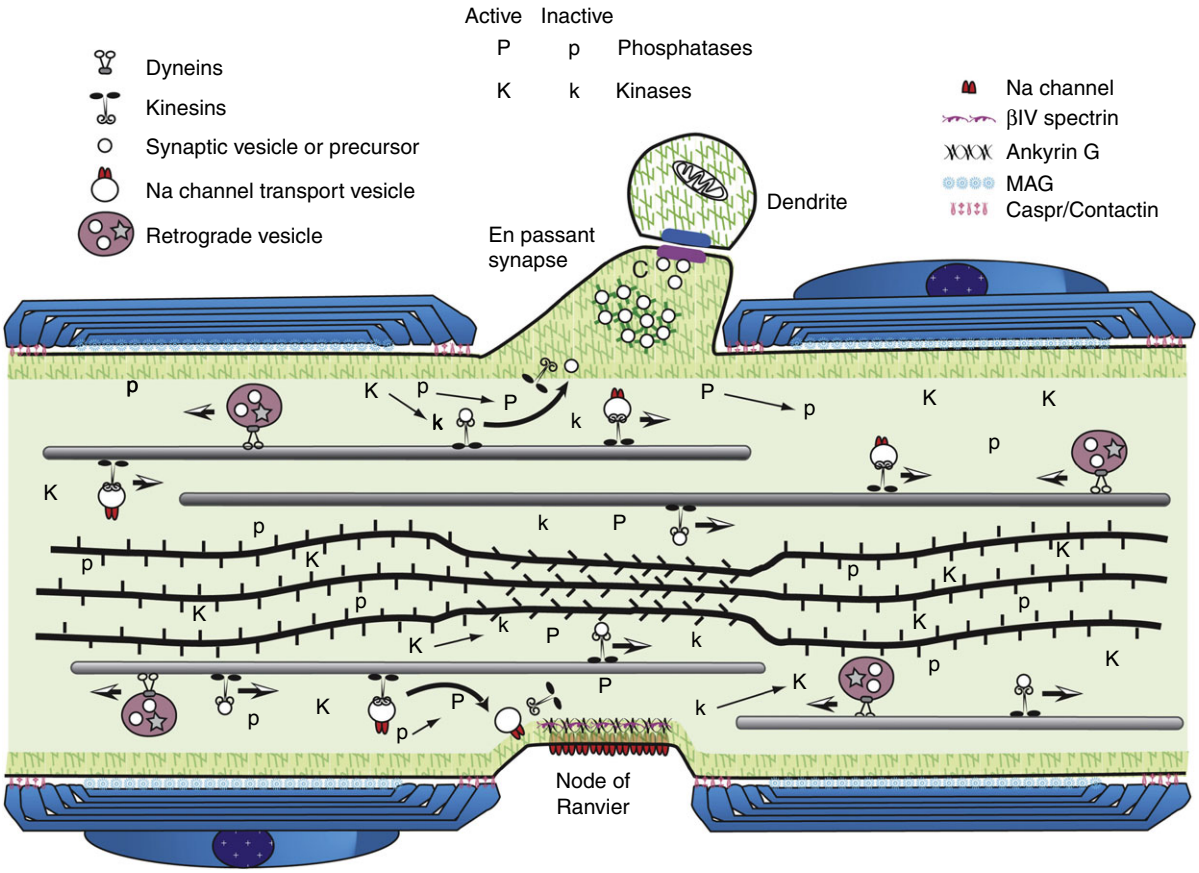


FIGURE 4.9 Axonal dynamics in a myelinated axon from the peripheral nervous system (PNS). Axons are in a constant flux with many concurrent dynamic processes. This diagram illustrates a few of the many dynamic events occurring at a node of Ranvier in a myelinated axon from the PNS. Axonal transport moves cytoskeletal structures, cytoplasmic proteins, and membrane-bound organelles from the cell body toward the periphery (from right to left). At the same time, other vesicles return to the cell body by retrograde transport (retrograde vesicle). Membrane-bound organelles are moved along microtubules by motor proteins such as the kinesins and cytoplasmic dyneins. Each class of organelles must be directed to the correct functional domain of the neuron. Synaptic vesicles must be delivered to a presynaptic terminal to maintain synaptic transmission. In contrast, organelles containing sodium channels must be targeted specifically to nodes of Ranvier for saltatory conduction to occur. Cytoskeletal transport is illustrated by microtubules (rods in the upper half of the axon) and neurofilaments (bundle of rope-like rods in the lower half of the axon) representing the cytoskeleton. They move in the anterograde direction as discrete elements and are degraded in the distal regions. Microtubules and neurofilaments interact with each other transiently during transport, but their distribution in axonal cross sections suggests that they are not stably cross-linked. In axonal segments without compact myelin, such as the node of Ranvier or following focal demyelination, a net dephosphorylation of neurofilament side arms allows the neurofilaments to pack more densely. Myelination is thought to alter the balance between kinase (K indicates an active kinase; k is an inactive kinase) and phosphatase (P indicates an active phophatase; p is an inactive phosphatase) activity in the axon. Most kinases and phosphatases have multiple substrates, suggesting a mechanism for targeting vesicle proteins to specific axonal domains. Local changes in the phosphoryation of axonal proteins may alter the binding properties of proteins. The action of synapsin I in squid axoplasm suggests that dephosphorylated synapsin cross-links synaptic vesicles to microfilaments. When a synaptic vesicle encounters the dephosphorylated synapsin and actin-rich matrix of a presynaptic terminal, the vesicle is trapped at the terminal by inhibition of further axonal transport, effectively targeting the synaptic vesicle to a presynaptic terminal. Similarly, a sodium channel-binding protein may be present at nodes of Ranvier in a high-affinity state (i.e., dephosphorylated). Transport vesicles for nodal sodium channels (Na channel vesicle) would be captured upon encountering this domain, effectively targeting sodium channels to the nodal membrane. Interactions between cells could in this manner establish the functional architecture of the neuron.

during stimulation. If nonphosphorylated synapsin were present at the border between the axon and the presynaptic terminal, then a fraction of the synaptic vesicles (or their precursors) traveling down the axon in fast axonal transport would be bound and cross-linked to the actin cytoskeleton that is enriched in the presynaptic terminal. These bound synaptic vesicles would become part of the reserve pool in the terminal at some distance from active zones. As the terminal was stimulated, local calmodulin-activated kinases would phosphorylate synapsin and allow these reserve synaptic vesicles to be mobilized. Consistent with this model, dephosphorylated synapsin has been shown to inhibit fast axonal transport when introduced directly into the axoplasm at concentrations comparable to those seen in the presynaptic terminal (McGuinness et al., 1989).

The role of kinases in this targeting appears to be a common feature in the regulation of axonal transport (Morfini, Szebenyi, Richards, & Brady, 2001). Local activation of kinases satisfies several criteria that any mechanism for targeting to specific neuronal subdomains must address. Specifically, the mechanism must be spatially restricted because distances to the cell body can be quite large. There must be some means to connect the targeting signal to an external microenvironment, such as an appropriate target cell. Finally, there must be a way of distinguishing subdomains so that synaptic vesicles are not delivered to nodes of Ranvier and voltage-gated sodium channels are not all targeted to the presynaptic terminal. The careful segregation of different organelles and polypeptides to different regions within a neuron suggests that highly efficient targeting mechanisms do exist.

Summary

A well-studied feature of the neuron is the phenomenon of axonal transport, which has been described in both anterograde and retrograde directions. The axonal transport system is responsible for the delivery of materials from the cell body to distant parts of the neuron, for membrane retrieval and circulation, and for uptake of materials from presynaptic terminals and dendrites and their delivery to the cell soma. The precise molecular mechanisms by which anterograde and retrograde transport can be targeted within an individual dendrite and/or axon are not understood at present.

Neurons and glial cells may have unusually large cell volumes enclosed within extensive plasma membrane surfaces. Nature has evolved a number of "universal" mechanisms in other systems and adapted them for the special needs of nervous tissue cells. The synthesis and delivery of components, and in particular proteins, to cytoplasmic organelles and cell surface subdomains engage general and evolutionarily conserved molecular mechanisms and pathways that are employed in single-cell yeasts as well as in cells in complex nervous tissue.

Once synthesized and sorted, most intracellular organelles (vesicles destined for axonal or dendritic domains, mitochondria, cytoskeletal components) must be distributed, and targeted, to precise intracellular locations. Because they are so extended in space, neurons and glial cells have adapted and developed to a high degree common mechanisms that operate to distribute components within all cells. In the neuron, movement of materials within the axon has been the central focus of most studies. The phenomenon of axoplasmic transport is now in some measure understood at the molecular level.

Dedication

We are very sad to report that David Colman died suddenly during the preparation of this edition. As a major contributor to our current understanding of myelination, David was a highly respected scientific colleague and, most of all, a good friend. We dedicate this chapter to his memory.

References

Allan, V. J. (2011). Cytoplasmic dynein. *Biochemical Society Transactions, 39,* 1169–1178.

Arnold, D. B. (2009). Actin and microtubule-based cytoskeletal cues direct polarized targeting of proteins in neurons. *Science Signaling, 2,* 49.

Baas, P. W., & Buster, D. W. (2004). Slow axonal transport and the genesis of neuronal morphology. *Journal of Neurobiology, 58,* 3–17.

Berl, S., Puszkin, S., & Nicklas, W. J. K. (1973). Actomyosin-like protein in brain. *Science, 179,* 441–446.

Brady, S. (1985). Axonal transport methods and applications. In A. Boulton, & G. Baker (Eds.), *Neuromethods, general neurochemical techniques* (pp. 419–476). Clifton, NJ: Humana Press.

Brady, S. T. (1991). Molecular motors in the nervous system. *Neuron, 7,* 521–533.

Brady, S. T. (1993). Axonal dynamics and regeneration. In A. Gorio (Ed.), *Neuroregeneration* (pp. 7–36). New York City, NY: Raven Press.

Brady, S. T., & Sperry, A. O. (1995). Biochemical and functional diversity of microtubule motors in the nervous system. *Current Opinion in Neurobiology, 5,* 551–558.

Brady, S. T., & Tai, L. (2012). Cell biology of the nervous system. In S. T. Brady, G. Siegel, R. W. Albers, & D. Price (Eds.), *Basic neurochemistry: Principles of molecular, cellular and medical neurobiology* (pp. 3–25). Boston, MA: Elseiver.

Braulke, T., & Bonifacino, J. S. (2009). Sorting of lysosomal proteins. *Biochimica et Biophysica Acta, 1793,* 605–614.

Brown, A. (2003). Live-cell imaging of slow axonal transport in cultured neurons. *Methods in Cell Biology, 71,* 305–323.

Colman, D. R., Kreibich, G., Frey, A. B., & Sabatini, D. D. (1982). Synthesis and incorporation of myelin polypeptides into CNS myelin. *Journal of Cell Biology, 95,* 598–608.

Doherty, G. J., & McMahon, H. T. (2009). Mechanisms of endocytosis. *Annual Review of Biochemistry, 78,* 857–902.

Dudek, J., Benedix, J., Cappel, S., Greiner, M., Jalal, C., Muller, L., & Zimmermann, R. (2009). Functions and pathologies of BiP and its interaction partners. *Cellular and Molecular Life Sciences, 66,* 1556–1569.

Endo, T., & Yamano, K. (2009). Multiple pathways for mitochondrial protein traffic. *Biological chemistry, 390,* 723–730.

Foth, B. J., Goedecke, M. C., & Soldati, D. (2006). New insights into myosin evolution and classification. *Proceedings of the National Academy of Sciences of the United States of America, 103,* 3681–3686.

Glick, B. S., & Nakano, A. (2009). Membrane traffic within the Golgi apparatus. *Annual Review of Cell and Developemental Biology, 25,* 113–132.

Grubb, M. S., & Burrone, J. (2010). Building and maintaining the axon initial segment. *Current Opinion in Neurobiology, 20,* 481–488.

Herrmann, H., Bar, H., Kreplak, L., Strelkov, S. V., & Aebi, U. (2007). Intermediate filaments: From cell architecture to nanomechanics. *Nature Reviews Molecular Cell Biology, 8,* 562–573.

Hirokawa, N., Noda, Y., Tanaka, Y., & Niwa, S. (2009). Kinesin superfamily motor proteins and intracellular transport. *Nature Reviews Molecular Cell Biology, 10,* 682–696.

Hirst, J., Barlow, L. D., Francisco, G. C., Sahlender, D. A., Seaman, M. N., Dacks, J. B., & Robinson, M. S. (2011). The fifth adaptor protein complex. *PLoS biology, 9,* e1001170.

Hochstrasser, M. (1995). Ubiquitin, proteasomes, and the regulation of intracellular protein degradation. *Current Opinion in Cell Biology, 7,* 215–223.

Kornfeld, R., & Kornfeld, S. (1985). Assembly of asparagine-linked oligosaccharides. *Annual Review of Biochemistry, 54,* 631–664.

Lasek, R. J., & Brady, S. T. (1982). The Axon: A prototype for studying expressional cytoplasm. In *Organization of the cytoplasm* (pp. 113–124). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Letourneau, P. C. (2009). Actin in axons: Stable scaffolds and dynamic filaments. *Results and Problems in Cell Differentiation, 48,* 65–90.

Ma, C., Agrawal, G., & Subramani, S. (2011). Peroxisome assembly: Matrix and membrane protein biogenesis. *The Journal of Cell Biology, 193,* 7–16.

Martin, K. C., & Zukin, R. S. (2006). RNA trafficking and local protein synthesis in dendrites: An overview. *Journal of Neuroscience, 26,* 7131–7134.

McGuinness, T. L., Brady, S. T., Gruner, J., Sugimori, M., Llinas, R., & Greengard, P. (1989). Phosphorylation-dependent inhibition by synapsin I of organelle movement in squid axoplasm. *Journal of Neuroscience, 9,* 4138–4149.

Morfini, G., Szebenyi, G., Richards, B., & Brady, S. T. (2001). Regulation of kinesin: Implications for neuronal development. *Developmental Neuroscience, 23,* 364–376.

Palade, G. (1975). Intracellular aspects of the process of protein synthesis. *Science, 189,* 347–358.

Palay, S. L., Sotelo, C., Peters, A., and Orkand, P. M. (1968). The axon hillock and the initial segment. *Journal of Cell Biology, 38,* 193–201.

Peters, A., Palay, S. L., & Webster, H. D. (1991). *The fine structure of the nervous system: Neurons and their supporting cells* (3rd ed.). New York, NY: Oxford University Press.

Pigino, G., Song, Y., Kirkpatrick, L. L., & Brady, S. T. (2012). The cytoskeleton of neurons and glia. In S. T. Brady, G. Siegel, R. W. Albers, & D. Price (Eds.), *Basic neurochemistry: Principles of molecular, cellular and medical neurobiology* (pp. 101–118). Boston, MA: Elsevier.

Rapoport, T. A. (2007). Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature, 450,* 663–669.

Roll-Mecak, A., & McNally, F. J. (2010). Microtubule-severing enzymes. *Current Opinion in Cell Biology, 22,* 96–103.

Roy, S., Winton, M. J., Black, M. M., Trojanowski, J. Q., & Lee, V. M. (2007). Rapid and intermittent cotransport of slow component-b proteins. *Journal of Neuroscience, 27,* 3131–3138.

Sherman, D. L., and Brophy, P. J. (2005). Mechanisms of axon ensheathment and myelin growth. *Nature Reviews Neuroscience, 6,* 683–690.

Sudhof, T. C. (2004). The synaptic vesicle cycle. *Annual Review of Neuroscience, 27,* 509–547.

Susuki, K., & Rasband, M. N. (2008). Molecular mechanisms of node of Ranvier formation. *Current Opinion in Cell Biology, 20,* 616–623.

Tanaka, J., & Sobue, K. (1994). Localization and characterization of gelsolin in nervous tissues: Gelsolin is specifically enriched in myelin-forming cells. *Journal of Neuroscience, 14,* 1038–1052.

Tsukita, S., & Ishikawa, H. (1980). The movement of membranous organelles in axons. Electron microscopic identification of anterogradely and retrogradely transported organelles. *Journal of Cell Biology, 84,* 513–530.

Vouyiouklis, D. A., & Brophy, P. J. (1995). Microtubule-associated proteins in developing oligodendrocytes: transient expression of a MAP2c isoform in oligodendrocyte precursors. *Journal of Neuroscience Research, 42,* 803–817.

Walter, P., & Johnson, A. E. (1994). Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. *Annual Review of Cell Biology, 10,* 87–119.

Xu, T., & Xu, P. (2008). Searching for molecular players differentially involved in neurotransmitter and neuropeptide release. *Neurochemical Research, 33,* 1915–1919.

Zerial, M., & McBride, H. (2001). Rab proteins as membrane organizers. *Nature Reviews Molecular Cell Biology, 2,* 107–117.

Scott T. Brady, David R. Colman, and Peter J. Brophy

CHAPTER

5

Membrane Potential and Action Potential

The communication of information between neurons and between neurons and muscles or peripheral organs requires that signals travel over considerable distances. A number of notable scientists have contemplated the nature of this communication through the ages. In the second century A.D., the great Greek physician Claudius Galen proposed that "humors" flowed from the brain to the muscles along hollow nerves. A true electrophysiological understanding of nerve and muscle, however, depended on the discovery and understanding of electricity itself. The precise nature of nerve and muscle action became clearer with the advent of new experimental techniques by a number of European scientists, including Luigi Galvini, Emil Du Bois-Reymond, Carlo Matteucci, and Hermann von Helmholtz, to name a few (McComas, 2011). Through the application of electrical stimulation to nerves and muscles, these early electrophysiologists demonstrated that the conduction of commands from the brain to the muscle for the generation of movement was mediated by the flow of electricity along nerve fibers.

With the advancement of electrophysiological techniques, electrical activity recorded from nerves revealed that the conduction of information along the axon was mediated by the active generation of an electrical potential, called the action potential. But what precisely was the nature of these action potentials? To know this in detail required not only a preparation from which to obtain intracellular recordings but also one that could survive *in vitro*. The squid giant axon provided precisely such a preparation. Many invertebrates contain unusually large axons for the generation of escape reflexes; large axons conduct more quickly than small ones and so the response time for escape is reduced. The squid possesses an axon approximately 0.5 mm in diameter, large enough to be impaled by even a course micropipette (Fig. 5.1). By inserting a glass micropipette filled with a salt solution into the squid giant axon, Alan Hodgkin and Andrew Huxley demonstrated in 1939 that axons at rest are electrically polarized, exhibiting a resting membrane potential of approximately −60 mV inside versus outside. In the generation of an action potential, the polarization of the membrane is removed (referred to as depolarization) and exhibits a rapid swing toward, and even past, 0 mV (Fig. 5.1). This depolarization is followed by a rapid swing in the membrane potential to more negative values, a process referred to as hyperpolarization. The membrane potential following an action potential typically becomes even more negative than the original value of approximately −60 mV. This period of increased polarization is referred to as the after-hyperpolarization or the undershoot.

The development of electrophysiological techniques to the point that intracellular recordings could be obtained from the small cells of the mammalian nervous system revealed that action potentials in these neurons are generated through mechanisms similar to that of the squid giant axon.

It is now known that action potential generation in nearly all types of neurons and muscle cells is accomplished through mechanisms similar to those first detailed in the squid giant axon by Hodgkin and Huxley. This chapter considers the cellular mechanisms by which neurons and axons generate a resting membrane potential, how signals passively propagate through neuronal processes, and finally, how active propagation of electrical signals is achieved through generation of the action potential.

© 2013 Elsevier Inc. All rights reserved.

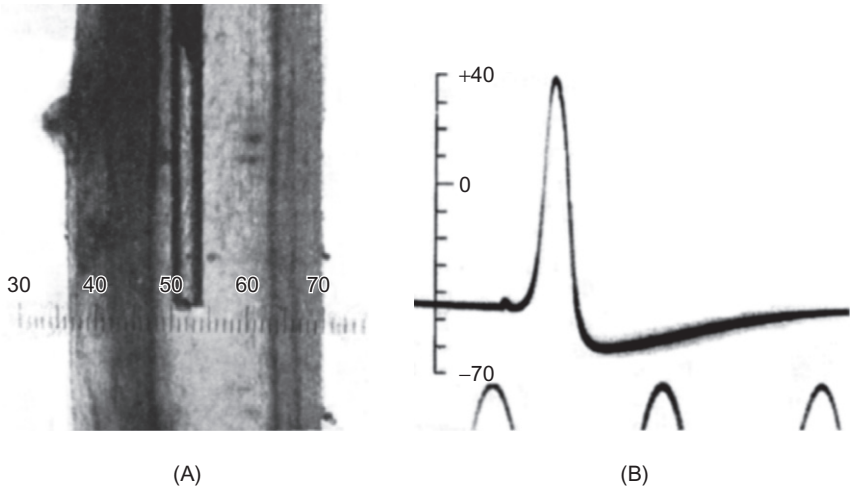


FIGURE 5.1 Intracellular recording of the membrane potential and action potential generation in the squid giant axon. (A) A glass micropipette, about 100 μm in diameter, was filled with seawater and lowered into the giant axon of the squid after it had been dissected free. The axon is about 1 mm in diameter and is transilluminated from behind. (B) One action potential recorded between the inside and the outside of the axon. Peaks of a sine wave at the bottom provided a scale for timing, with 2 ms between peaks. *From Hodgkin and Huxley (1939).*

MEMBRANE POTENTIAL

Membrane Potential Is Generated by the Differential Distribution of Ions

Through the operation of ionic pumps and special ionic buffering mechanisms, neurons actively maintain precise internal concentrations of several important ions, including Na^+, K^+, Cl^-, and Ca^{2+}. The mechanisms by which they do so are illustrated in Figures 5.2 and 5.3. The intracellular and extracellular concentrations of Na^+, K^+, Cl^-, and Ca^{2+} differ markedly (Fig. 5.2); K^+ is actively concentrated inside the cell, and Na^+, Cl^-, and Ca^{2+} are actively extruded to the extracellular space. However, this does not mean that the cell is filled only with positive charge; anions to which the plasma membrane is impermeant are also present inside the cell and almost balance the high concentration of K^+. The osmolarity inside the cell is approximately equal to that outside the cell.

Electrical and Thermodynamic Forces Determine the Passive Distribution of Ions

Ions tend to move down their concentration gradients through specialized ionic pores, known as ionic channels, in the plasma membrane. Through simple laws of thermodynamics, the high concentration of K^+ inside glial cells, neurons, and axons results in a tendency for K^+ ions to diffuse down their concentration gradient and leave the cell or cell process (Fig. 5.3). However, the movement of ions across the membrane also results in a redistribution of electrical charge. As K^+ ions move down their concentration gradient, the intracellular voltage becomes more negative, and this increased negativity results in an electrical attraction between the negative potential inside the cell and the positively charged, K^+ ions, thus offsetting the outward flow of these ions. The membrane is selectively permeable—that is, it is impermeable to the large anions inside the cell, which cannot follow the potassium ions across the membrane. At some membrane potential, the "force" of the electrostatic attraction between the negative membrane potential inside the cell and the positively charged K^+ ions will exactly balance the thermal "forces" by which K^+ ions tend to flow down their concentration gradient (Fig. 5.3). In this circumstance, it is equally likely that a K^+ ion exits the cell by movement down the concentration gradient as it is that a K^+ ion enters the cell due to the attraction between the negative membrane potential and the positive charge of this ion. At this membrane potential, there is no net flow of K^+ (the same number of K^+ ions enter the cell as leave the cell per unit time) and these ions are said to be in equilibrium. The membrane potential at which this occurs is known as the equilibrium potential. (See Box 5.1 for calculation of the equilibrium potential.)

To illustrate, let us consider the passive distribution of K^+ ions in the squid giant axon as studied by Hodgkin and Huxley. The K^+ concentration $[K^+]$ inside the squid giant axon is about 400 mM, whereas the $[K^+]$ outside the axon is about 20 mM. Because $[K^+]_i$ is greater than $[K^+]_o$, potassium ions will tend to flow down their concentration gradient, from inside to outside the axon, taking positive charge with them. The equilibrium potential (at which the tendency for K^+ ions to flow down their concentration gradient will be exactly offset by the attraction for K^+ ions to enter the cell because of

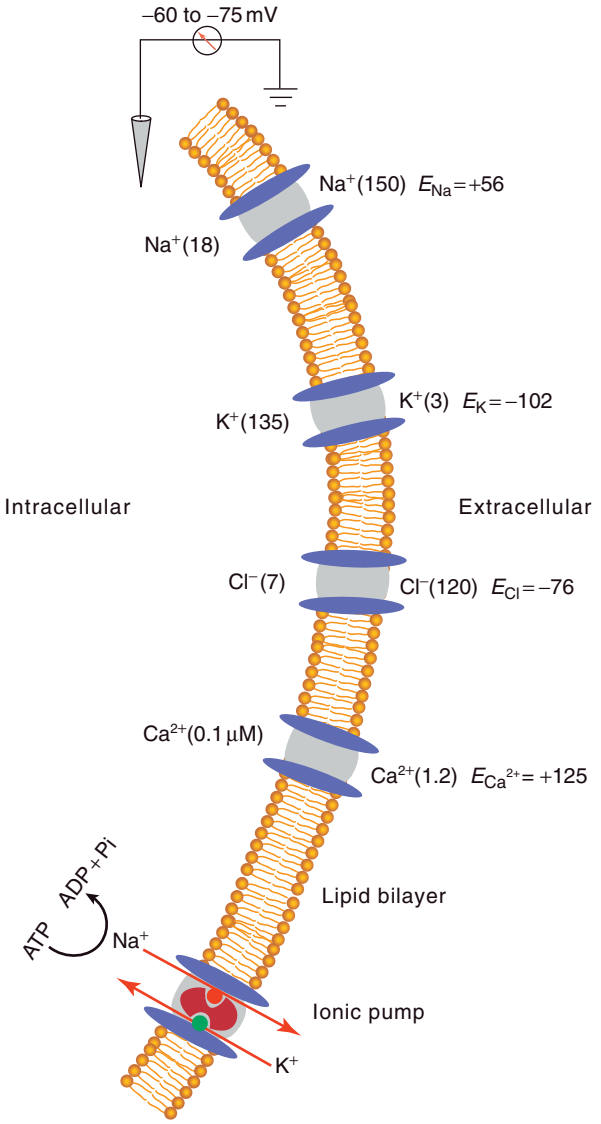


FIGURE 5.2 Differential distribution of ions inside and outside plasma membrane of neurons and neuronal processes, showing ionic channels for Na^+, K^+, Cl^-, and Ca^{2+}, as well as an electrogenic Na^+–K^+ ionic pump (also known as Na^+, K^+-ATPase). Concentrations (in millimoles except that for intracellular Ca^{2+}) of the ions are given in parentheses; their equilibrium potentials (E) for a typical mammalian neuron are indicated.

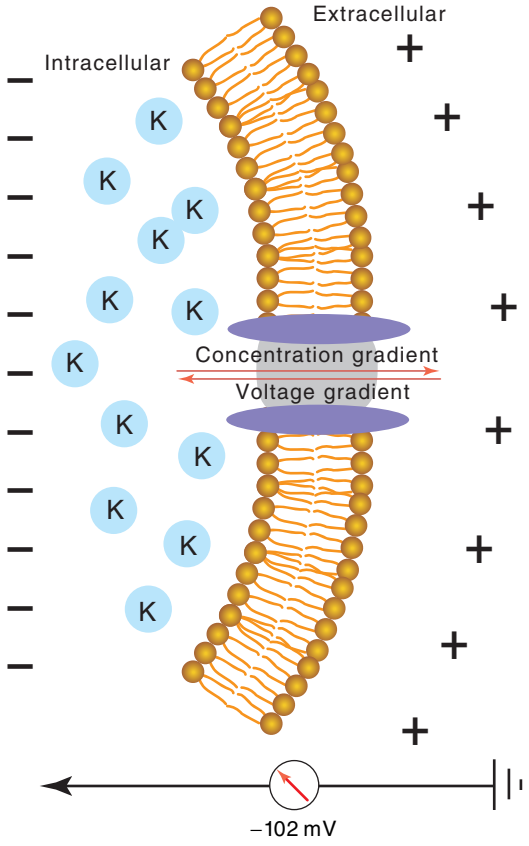


FIGURE 5.3 The equilibrium potential is influenced by the concentration gradient and the voltage difference across the membrane. Neurons actively concentrate K^+ inside the cell. These K^+ ions tend to flow down their concentration gradient from inside to outside the cell. However, the negative membrane potential inside the cell provides an attraction for K^+ ions to enter or remain within the cell. These two factors balance one another at the equilibrium potential, which in a typical mammalian neuron is –102 mV for K^+.

the negative charge inside the cell) at a room temperature of 20°C can be calculated by the Nernst equation as such:

$$E_K = 58.2 \log_{10}(20/400) = -76\ \text{mV}.$$

Therefore, at a membrane potential of –76 mV, K^+ ions have an equal tendency to flow either into or out of the axon. The concentrations of K^+ in mammalian neurons and glial cells differ considerably from that in the squid giant axon, which is adapted to live in seawater. By substituting 3.1 mM for $[K^+]_o$ and 140 mM for $[K^+]_i$ in the Nernst equation, with mammalian body temperature, T = 37°C, we obtain

$$E_K = 61.5 \log_{10}(3.1/140) = -102\ \text{mV}.$$

Movements of Ions Can Cause Either Hyperpolarization or Depolarization

In mammalian cells, at membrane potentials positive to –102 mV, K^+ ions tend to flow out of the cell. Increasing the ability of K^+ ions to flow across the membrane—that is, increasing the conductance of the membrane to K^+ (gK) causes the membrane potential to become more negative, or hyperpolarized, due to the exiting of positively charged ions from inside the cell (Fig. 5.4).

At membrane potentials negative to –102 mV, K^+ ions tend to flow into the cell; increasing the membrane conductance to K^+ causes the membrane potential to become more positive, or depolarized, due to the flow of positive charge into the cell. The membrane potential

BOX 5.1

NERNST EQUATION

The equilibrium potential is determined by (1) the concentration of the ion inside and outside the cell, (2) the temperature of the solution, (3) the valence of the ion, and (4) the amount of work required to separate a given quantity of charge. The equation that describes the equilibrium potential was formulated by a German physical chemist named Walter Nernst in 1888:

$$E_{ion} = RT/zF \cdot \ln([ion]_o/[ion]_i)$$

Here, E_{ion} is the membrane potential at which the ionic species is at equilibrium, R is the gas constant [8.315 J per Kelvin per mole (J K^{-1} mol^{-1})], T is the temperature in Kelvins ($T_{Kelvin} = 273.16 + T_{Celcius}$), F is Faraday's constant [96,485 coulombs per mole (C mol^{-1})], z is the valence of the ion, and $[ion]_o$ and $[ion]_i$ are the concentrations of the ion outside and inside the cell, respectively. For a monovalent, positively charged ion (cation) at room temperature (20°C), substituting the appropriate numbers and converting natural log (ln) into log base 10 ($\log_{10}$) result in

$$E_{ion} = 58.2 \log_{10}([ion]_o/[ion]_i)$$

at a body temperature of 37°C, the Nernst equation is

$$E_{ion} = 61.5 \log_{10}([ion]_o/[ion]_i).$$

David A. McCormick

at which the net current "flips" direction is referred to as the reversal potential. If the channels conduct only one type of ion (e.g., K^+ ions), then the reversal potential and the Nernst equilibrium potential for that ion coincide (Fig. 5.4A). Increasing the membrane conductance to K^+ ions while the membrane potential is at the equilibrium potential for K^+ (E_K) does not change the membrane potential because no net driving force causes K^+ ions to either exit or enter the cell. However, this increase in membrane conductance to K^+ decreases the ability of other species of ions to change the membrane potential because any deviation of the potential from E_K increases the drive for K^+ ions to either exit or enter the cell, thereby drawing the membrane potential back toward E_K (Fig. 5.4B). This effect is known as a "shunt" and is important for some effects of inhibitory synaptic transmission.

The exiting and entering of the cell by K^+ ions during generation of the membrane potential give rise to a curious problem. When K^+ ions leave the cell to generate a membrane potential, the concentration of K^+ changes both inside and outside the cell. Why does this change in concentration not alter the equilibrium potential, thus changing the tendency for K^+ ions to flow down their concentration gradient? The reason is that the number of K^+ ions required to leave the cell to achieve the equilibrium potential is quite small. For example, if a cell were at 0 mV and the membrane suddenly became permeable to K^+ ions, only about 10^{-12} mol of K^+ ions per square centimeter of membrane would move from inside to outside the cell in bringing the membrane potential to the equilibrium potential for K^+. In a spherical cell of 25 μm diameter, this would amount to an average decrease in intracellular K^+ of only about 4 μM (e.g., from 140 to 139.996 mM). However, there are instances when significant changes in the concentrations of K^+ may occur, particularly during the generation of pronounced activity, such as an epileptic seizure. During the occurrence of a tonic–clonic generalized (grand mal) seizure, large numbers of neurons discharge throughout the cerebral cortex in a synchronized manner. This synchronous discharge of large numbers of neurons significantly increases the extracellular K^+ concentration, by as much as a couple of millimoles, resulting in a commensurate positive shift in the equilibrium potential for K^+. This shift in the equilibrium potential can increase the excitability of affected neurons and neuronal processes and thus promote the spread of the seizure activity. Fortunately, the extracellular concentration of K^+ is tightly regulated and is kept at normal levels through uptake by glial cells, as well as by diffusion through the fluid of the extracellular space.

As is true for K^+ ions, each of the membrane-permeable species of ions possesses an equilibrium potential that depends on the concentration of that ion inside and outside the cell. Thus, equilibrium potentials may vary between different cell types, such as those found in animals adapted to live in salt water versus mammalian neurons. In mammalian neurons, the equilibrium potential is approximately 56 mV for Na^+, approximately −76 mV for Cl^-, and about 125 mV for Ca^{2+} (see Fig. 5.2). Thus, increasing the membrane conductance to Na^+ (gNa) through the opening of Na^+ channels depolarizes the membrane potential toward 56 mV; increasing the membrane conductance to Cl^- brings the membrane potential closer to −76 mV; and finally increasing the membrane conductance to Ca^{2+} depolarizes the cell toward 125 mV.

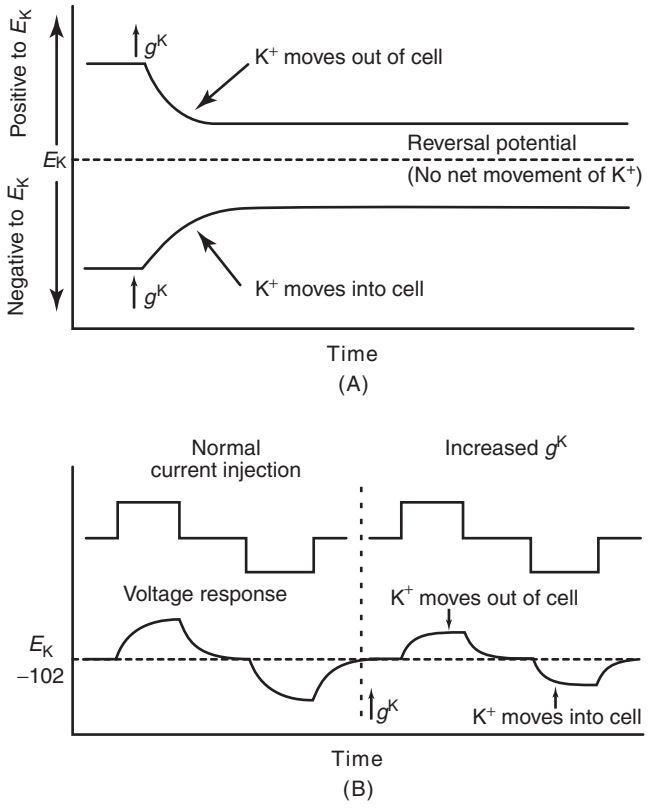


FIGURE 5.4 Increases in K^+ conductance can result in hyperpolarization, depolarization, or no change in membrane potential. (A) Opening K^+ channels increases the conductance of the membrane to K^+, denoted gK. If the membrane potential is positive to the equilibrium potential (also known as the reversal potential) for K^+, then increasing gK will cause some K^+ ions to leave the cell, and the cell will become hyperpolarized. If the membrane potential is negative to E_K when gK is increased, then K^+ ions will enter the cell, therefore making the inside more positive (more depolarized). If the membrane potential is exactly E_K when gK is increased, then there will be no net movement of K^+ ions. (B) Opening K^+ channels when the membrane potential is at E_K does not change the membrane potential; however, it reduces the ability of other ionic currents to move the membrane potential away from E_K. For example, a comparison of the ability of the injection of two pulses of current, one depolarizing and one hyperpolarizing, to change the membrane potential before and after opening K^+ channels reveals that increases in gK decrease the responses of the cell noticeably.

Na^+, K^+, and Cl^- Contribute to the Determination of the Resting Membrane Potential

If a membrane is permeable to only one ion and no electrogenic ionic pumps are operating (see next section), then the membrane potential is necessarily at the equilibrium potential for that ion. At rest, the plasma membrane of most cell types is not at the equilibrium potential for K^+ ions, indicating that the membrane is also permeable to other types of ions. For example, the resting membrane of the squid giant axon is permeable to Cl^- and Na^+, as well as K^+, due to the presence of ionic channels that not only allow these ions to pass but also are open at the resting membrane potential. Because the membrane is permeable to K^+, Cl^-, and Na^+, the resting potential of the squid giant axon is not equal to E_K, E_{Na}, or E_{Cl}, but is somewhere in between these three. A membrane permeable to more than one ion has a steady-state membrane potential whose value is between those of the equilibrium potentials for each of the permeant ions (Box 5.2).

Different Types of Neurons Have Different Resting Potentials

Intracellular recordings from neurons in the mammalian central nervous system (CNS) reveal that different types of neurons exhibit different resting membrane potentials. Indeed, some types of neurons do not even exhibit a true "resting" membrane potential; they spontaneously and continuously generate action potentials even in the total lack of synaptic input. In the visual system, intracellular recordings have shown that photoreceptor cells of the retina—the rods and cones—have a membrane potential of approximately −40 mV at rest and are hyperpolarized when activated by light. Cells in the dorsal lateral geniculate nucleus, which receive axonal input from the retina and project to the visual cortex, have a resting membrane potential of approximately −70 mV during sleep and −55 mV during waking, whereas pyramidal neurons of the visual cortex have a resting membrane potential of about −75 mV. Presumably, the resting membrane potentials of different cell types in the central and peripheral nervous system are highly regulated and are functionally important. For example, the depolarized membrane potential of photoreceptors presumably allows the membrane potential to move in both negative and positive directions in response to changes in light intensity. The hyper-polarized membrane potential of thalamic neurons during sleep (−70 mV) dramatically decreases the flow of information from the sensory periphery to the cerebral cortex, presumably to allow the cortex to be relatively undisturbed during sleep, and the 20 mV membrane potential between the resting potential and the action potential threshold in cortical pyramidal cells allows these cells to be strongly influenced by subthreshold barrages of synaptic potentials from other cortical neurons, which provide the context under which we think and behave (see Chapter 9).

Ionic Pumps Actively Maintain Ionic Gradients

Because the resting membrane potential of a neuron is not at the equilibrium potential for any particular ion, ions constantly flow down their concentration gradients. This flux becomes considerably larger with the

BOX 5.2

GOLDMAN-HODGKIN-KATZ EQUATION

An equation developed by Goldman and later used by Alan Hodgkin and Bernard Katz describes the steady-state membrane potential for a given set of ionic concentrations inside and outside the cell and the relative permeabilities of the membrane to each of those ions:

$$V_m \frac{RT}{F} \cdot \ln\left(\frac{(p_K[K^+]_o + p_{Na}[Na^+]_o + p_{Cl}[Cl^-]_i)}{(p_K[K^+]_i + p_{Na}[Na^+]_i + p_{Cl}[Cl^-]_o)}\right).$$

The relative contribution of each ion is determined by its concentration differences across the membrane and the relative permeability (p_K, p_{Na}, p_{Cl}) of the membrane to each type of ion. If a membrane is permeable to only one ion, then the Goldman–Hodgkin–Katz equation reduces to the Nernst equation. In the squid giant axon, at resting membrane potential, the permeability ratios are

$$p_K : p_{Na} : p_{Cl} = 1.00 : 0.04 : 0.45.$$

The membrane of the squid giant axon, at rest, is most permeable to K^+ ions, less so to Cl^-, and least permeable to Na^+. (Chloride appears to contribute considerably less to the determination of the resting potential of mammalian neurons.) These results indicate that the resting membrane potential is determined by the resting permeability of the membrane to K^+, Na^+, and Cl^-. In theory, this resting membrane potential may be anywhere between E_K (e.g., −76 mV) and E_{Na} (55 mV). For the three ions at 20°C, the equation is

$$V_m = \frac{58.2 \log_{10}\{(1 \cdot 20 + 0.04 \cdot 440 + 0.45 \cdot 40)}{(1 \cdot 400 + 0.04 \cdot 50 + 0.45 \cdot 560)\}}$$
$$= -62 \text{ mV}.$$

This suggests that the squid giant axon should have a resting membrane potential of −62 mV. In fact, the resting membrane potential may be a few millivolts hyperpolarized to this value through the operation of the electrogenic Na^+–K^+ pump.

David A. McCormick

generation of electrical and synaptic potentials because ionic channels are opened by these events. Although the absolute number of ions traversing the plasma membrane during each action potential or synaptic potential may be small in individual cells, the collective influence of a large neural network of cells, such as in the brain, and the presence of ion fluxes even at rest can substantially change the distribution of ions inside and outside neurons. Cells have solved this problem with the use of active transport of ions against their concentration gradients. The proteins that actively transport ions are referred to as ionic pumps, of which the Na^+–K^+ pump is perhaps the most thoroughly understood (Lauger, 1991). The Na^+–K^+ pump is stimulated by increases in the intracellular concentration of Na^+ and moves Na^+ out of the cell while moving K^+ into it, achieving this task through the hydrolysis of ATP (Fig. 5.2). Three Na^+ ions are extruded for every two K^+ ions transported into the cell. Because of the unequal transport of ions, the operation of this pump generates a hyper-polarizing electrical potential and is said to be electrogenic. The Na^+–K^+ pump typically results in the membrane potential of the cell being a few millivolts more negative than it would be otherwise.

The Na^+–K^+ pump consists of two subunits, α and β, arranged in a tetramer $(\alpha\beta)_2$. The Na^+–K^+ pump is believed to operate through conformational changes that alternatively bind Na^+ ions from the inside of the cell and release them to the extracellular space, while binding K^+ ions from outside the cell and releasing them into the intracellular mileau (Fig. 5.2). Such a conformation change is due to the phosphorylation and dephosphorylation of the protein.

The membranes of neurons and glia contain multiple types of ionic pumps, used to maintain the proper distribution of each ionic species important for cellular signaling (Reithmeier, 1994). Many of these pumps are operated by the Na^+ gradient across the cell, whereas others operate through a mechanism similar to that of the Na^+–K^+ pump (i.e., the hydrolysis of ATP). For example, the calcium concentration inside neurons is kept to very low levels (typically 50–100 nM) through the operation of both types of ionic pumps, as well as special intracellular Ca^{2+} buffering mechanisms. Ca^{2+} is extruded from neurons through both a Ca^{2+}, Mg^{2+}-ATPase and a Na^+–Ca^{2+} exchanger. The Na^+–Ca^{2+} exchanger is driven by the Na^+ gradient across the membrane and extrudes one Ca^{2+} ion for each Na^+ ion allowed to enter the cell.

The Cl^- concentration in neurons is actively maintained at a low level through operation of a chloride-bicarbonate exchanger, which brings in one ion of Na^+ and one ion of HCO_3^- for each ion of Cl^- extruded. Intracellular pH can also markedly affect neuronal excitability and is therefore tightly regulated, in part by a Na^+–H^+ exchanger that extrudes one proton for each Na^+ allowed to enter the cell.

Summary

The membrane potential is generated by the unequal distribution of ions, particularly K^+, Na^+, and Cl^-, across the plasma membrane. This unequal distribution of ions is maintained by ionic pumps and exchangers. K^+ ions are concentrated inside the neuron and tend to flow down their concentration gradient, leading to a hyperpolarization of the cell. At the equilibrium potential, the tendency of K^+ ions to flow out of the cell will be exactly offset by the tendency of K^+ ions to enter the cell due to the attraction of the negative potential inside the cell. The resting membrane is also permeable to Na^+ and Cl^- and therefore the resting membrane potential is approximately −75 to −40 mV, in other words, substantially positive to E_K.

PASSIVE ELECTRICAL PROPERTIES OF THE NEURON

Electrical activity in the nervous system not only conducts actively along neuronal processes, owing to the opening and closing of ion channels, but also moves passively, with electrical charges pushing or pulling other charges both inside and outside thin neuronal elements. How electrical charge moves passively through neurons shapes how these cells process synaptic inputs and generate action potential outputs, and thus interact with other neurons. Three properties that strongly influence the passive response properties of neuronal elements are: membrane resistance, axial resistance, and membrane capacitance (Fig. 5.5A). These three properties not only determine the amplitude of the neuron's passive response to a synaptic input, but also the amplitude and speed of spread of this input to other parts of the neuron and the properties of the action potential response.

The outer membrane of the neuron, being composed of phosholipids and ion channels, offers resistance to the flow of charged ions and thus exhibits a specific membrane resistance, R_m. Conversely, the cell membrane can be considered to exhibit a specific level of conductivity (G_m; conductance) to electrical currents, which is mathematically described as $1/R_m$. The amount of current that flows across a given conductance is proportional to the voltage applied, in what is known as Ohm's law:

$$I = VG \quad \text{or} \quad I = V/R \quad \text{or} \quad V = IR$$

This is a linear relationship that states that the larger the voltage that is applied to a fixed conductance, the greater the current that will flow. Likewise, increasing membrane conductance (e.g., opening ion channels) increases the current that flows across the membrane in the presence of a fixed voltage. Imagine measuring the flow of water through a pipe. The rate of flow of the water (current) is proportional to the pressure applied (voltage) as well as the size of the pipe (conductance or resistance). Increasing the diameter of the pipe increases the rate at which water flows as long as the pressure doesn't change. The change in voltage (pressure) that occurs with a fixed current flow against a resistance can also be determined (V = IR). In a neuron, for example, the opening of enough ion channels to produce 100 pA of current in a cell that is 100 Megohms in resistance will result in a change in membrane potential of 10 mV at the point of entry of the ionic current and at steady state. The qualifiers "at the point of entry" and "at steady state" are given because neurons have two properties that strongly shape the response to ionic currents: Current flows down processes and not just across membranes, and membranes not only have resistance but also capacitance (Fig. 5.5A).

Capacitance is the ability to store electrical charge and occurs whenever two electrical conductors (e.g., cytoplasm and extracellular fluid) are separated by a nonconductor (e.g., lipid membrane). The gathering of ions near one side of the membrane will push like charges away, and attract opposite charges toward, the other side of the membrane. For example, the entry of positive charge into a neuron will result in an increase in positive ions under the membrane. These positive ions on the inside of the cell membrane will repel positive ions away from the outside of the membrane, resulting in a capacitative current (I_c). In a way, capacitors conduct electricity, just as conductors do. However, there are marked differences. While a resistor conducts a constant flow of current (e.g., charges/second) with a constant voltage (I = V/R), a capacitor requires an ever increasing voltage (e.g., volts/second) in order to deliver a constant flow of current. In other words, the ability of a capacitor to provide a conduit for current flow is dependent upon the rate of change of membrane potential; it is frequency dependent. In fact, one way of thinking of a capacitor is that it is a variable resistor in which the resistance depends on frequency: at high frequencies (i.e., volts/sec is high), the resistance is low and current flows readily, while at low frequencies, the resistance is high and current flow is limited. Let's consider why this is so and how this affects the manner in which neurons behave.

A capacitor is defined by the amount of charge (measured in coulombs; Q = 6.24×10^{18} electrons or protons) that are stored on the opposing plates per volt (C = Q/V). The current flowing through a capacitor (I_c), measured in amperes, is dQ/dt (coulombs/sec), which is equal to CdV/dt (i.e., the time differential of Q = CV). Thus, the greater the rate of change of membrane potential, the greater the flow of current through the membrane capacitance. Likewise, when the membrane potential is steady,

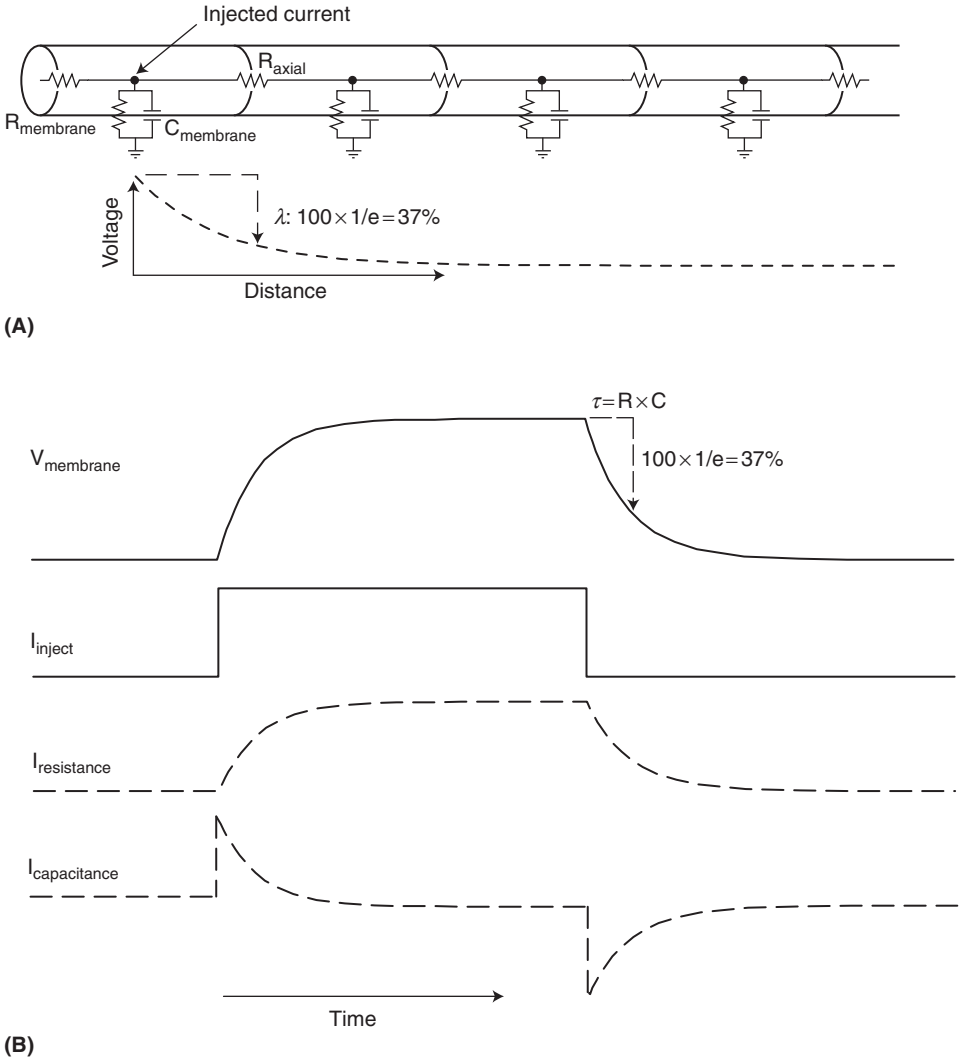


FIGURE 5.5 Passive response properties of neuronal element. (A) Idealized cylindrical neuronal process. The injection of current into a particular point in the process will result in the flow of ions across three paths: down the process through the axial resistance (R_{axial}) and across the membrane through the membrane resistance ($R_{membrane}$) and capacitance ($C_{membrane}$). The leak of current out of the process as it flows down the process and across the membrane will result in an exponential decrease in voltage. The point at which the voltage has reduced to 1/e (37%) of its original value is known as the length constant (λ). (B) Response of the membrane to the injection of a constant current pulse. Initially the injected current flows readily through the membrane capacitance (bottom trace). As this capacitative current ($I_{capacitance}$) decreases (bottom trace), there is a rise in charge inside the membrane and thus the voltage rises ($V_{membrane}$; top trace). This increase in voltage inside the cell results in an increase in current flow across the membrane resistance ($I_{resistance}$). Turning off the current pulse will result in the opposite flow of current through the capacitance, and the membrane potential will return to it's previous level with an exponential time course. The time it takes to move to 1/e (37%) of the steady state change in membrane potential is known as the membrane time constant (τ).

there is no current flow through the capacitance. A second big difference between current flow through capacitance versus resistance is that the capacitative current is made up of the movement of different ions on each side of the membrane; the gathering of ions on one side influences the distribution of ions on the other. In contrast, the movement of ions through the membrane resistance, which is made up largely of ionic pores, consists of the actual movement of ions across the membrane and therefore can result in changes in ion concentrations inside and outside the cell.

Membranes are characterized by a capacitance and resistance in parallel (Fig. 5.5). If we apply a steady current to the inside of the membrane, at first the capacitance acts as a low-resistance pathway and there is a large flow of current (initially equal to that of the applied current) through the membrane capacitance (Fig. 5.5B; $I_{capacitance}$). The change in distribution of charge across the membrane will alter the membrane potential, and this will provide a new driving force for ions to move through the ionic pores that make up the membrane resistance/conductance (Fig. 5.5B; $I_{resistance}$). As ions move through

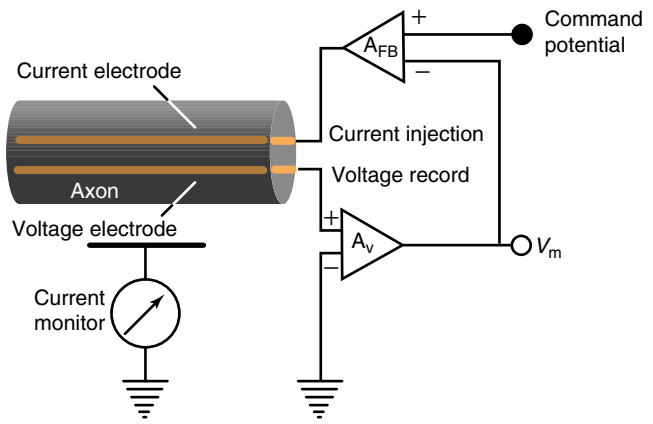


FIGURE 5.6 The voltage-clamp technique keeps the voltage across the membrane constant so that the amplitude and time course of ionic currents can be measured. In the two-electrode voltage-clamp technique, one electrode measures the voltage across the membrane while the other injects current into the cell to keep the voltage constant. The experimenter sets a voltage to which the axon or neuron is to be stepped (the command potential). Current is then injected into the cell in proportion to the difference between the present membrane potential and the command potential. This feedback cycle occurs continuously, thereby clamping the membrane potential to the command potential. By measuring the amount of current injected, the experimenter can determine the amplitude and time course of the ionic currents flowing across the membrane.

the membrane resistance/conductance, the rate at which they accumulate under the membrane will slow down, resulting in a decreased rate of change of membrane potential (dV/dt) and consequently a decrease in current flow through the membrane capacitance. Eventually, the membrane potential will settle at a value dictated by the magnitude of the current injected and the membrane resistance, according to Ohm's law.

The time course over which the change in membrane potential occurs in response to a step change in current is described by an exponential function in which the time constant τ_m, is determined by the product of the specific membrane resistance and membrane capacitance per unit area (τ_m = RC) and is equal to the time it takes for the membrane potential to reach 37% (100 * 1/e) of its steady-state value (Fig. 5.5B). Once the current step is turned off, current will then flow across the capacitor in the opposite direction. In the example of Figure 5.5B, the removal of the positive current step leaves the inner membrane with a buildup of positive charge (and the outer membrane with an excess of negative charge). As the inner positive and outer negative charges move away from the membrane, the membrane potential will decrease exponentially, and thus the current through the membrane resistance (conductance) will also decrease exponentially (Fig. 5.5B).

Let's consider the hydraulic analogy again. Imagine two parallel pipes connected together on both ends. One pipe contains a chamber with a diaphragm inside (a capacitor), and the other contains a constriction (a resistor). Applying a steady flow of water to one end of the two parallel connected pipes will at first cause an equal movement of water at the other end by distending the diaphragm (capacitor). However, as the diaphragm stretches, it will start to resist further movement, and thus the water pressure inside the pipes will begin to rise. This rise in pressure will force water through the parallel (resistor) pipe at a rate determined by the pressure/resistance. Since water is now flowing through the "resistor" pipe, there is less further accumulation of water against the diaphragm, resulting in a decreased rate of water flow through the "capacitor" path. Eventually, all of the water flow will be through the "resistive" pipe and none will be through further distension of the diaphragm. Once the steady flow of water is removed, however, the diaphragm will provide a temporary source for water to flow once again, resulting in the slow dissipation of the "capacitive current" and a slow return to the precurrent water pressure (e.g., Fig. 5.5B).

An important consideration in passive properties of membrane responses, owing in part to the parallel capacitance/resistance properties of the membrane, is their frequency filtering properties. In the example in Figure 5.5B, a square pulse of current was applied, yet the membrane responded with a gradual increase (at the onset of current) and then decrease (at the offset of current) in membrane potential; the high-frequency components of the current step were filtered out by the capacitive/resistive properties of the membrane. Neuronal membranes naturally respond to some frequencies (e.g., low frequencies) better than others (e.g., high frequencies). Since conductance through the membrane capacitance is frequency dependent, being highest with rapid changes in current, the membrane capacitor, in parallel with membrane resistance, acts as a low-pass filter for changes in membrane potential (voltage across the membrane). Higher frequencies are decreased in amplitude, while low frequencies are preserved (in other words, the high-frequency currents escape from the neuron passing easily through the capacitance of the membrane, thereby reducing the ability of high-frequency current entry into the neuron to build up voltage ["pressure"] inside the cell). For a neuron with a membrane time constant of 20 msec, the "cutoff" frequency (defined as $f_c = 1/(2\pi RC)$) of the filtering properties of the membrane will be 8 Hz, meaning that increasing frequency above 8 Hz will result in a larger and larger decrease in the membrane voltage response. One often overlooked property of filters is that they not only change the amplitude but also the phase (timing) of an input in a frequency-dependent manner.

Next let's consider how current flows along a neuronal process. The entry of ions into the cell at one point will result not only in the flow of current across the membrane

resistance and capacitance but also the movement of current away from the entry point along the neuronal process. This movement of charge along the inner part of the neuron is known as "axial current." The resistance to the movement of axial current, owing to the properties of cytoplasm and organelles within the neuron, is referred to as axial resistance (Fig. 5.5A). Axial resistance in neurons is usually small in comparison to membrane resistance, allowing current to flow readily down neuronal processes. However, as the current flows, some of it escapes through the local membrane resistance and capacitance of each neuronal compartment that it enters. Imagine a garden hose that not only is elastic (can bulge, representing capacitance) but also has small holes in it all along its length. The application of a given value of water pressure at the end of the hose will result in a gradient of pressure along the hose, owing to the local reduction in water pressure through the leaking and bulging of the hose. Likewise, in a neuronal process, the voltage (e.g., "pressure") decreases exponentially as the current spreads across the axial resistance and leaks out through the resistance and capacitance of the membrane (Fig. 5.5A). The point at which the membrane potential has fallen to 1/e (or 37%) of its original value is known as the length constant (λ) of the process (Fig. 5.5A). Neuronal processes such as dendrites and axons have characteristic length constants in the range of 0.1 to 1 mm at steady state, depending on their diameter. Larger-diameter processes, such as the squid giant axon, have lower axial resistance (higher axial conductance) and therefore a longer length constant, allowing electrical current to have a broader influence and to flow quickly. The length constant also depends upon the frequency of the input, since the outer membrane acts as a frequency dependent filter. This filtering property not only affects the amplitude of electrical events as they propagate, but also their phase. Since the dendritic processes of many neurons are within the 0.1–1 mm length constant, voltage changes at one part of a dendrite can passively spread to a significant portion of the local dendritic arbor. For the vast majority of axons, a length constant of 0.1–1 mm is insufficient to spread information to synaptic boutons, resulting in the need of action potentials to propagate information over long distances. It is interesting to note, however, that a length constant of 0.2–0.4 mm, even in the fine axons of the brain, suggests that local changes in membrane potential at the soma can actually propagate to nearby synaptic terminals and potentially influence synaptic transmission (Shu, Hasenstaub, Duque, Yu, & McCormick, 2006).

The flow of current in actual neurons is complex, owing to both the complicated geometries of dendritic and axonal arbors and the presence of ionic channels that actively open and close in response to changes in membrane voltage. These active properties of neurons result not only in the ability of the cell to communicate over long distances, but they also allow each cell type to exhibit its own unique electrophysiological properties that are useful for achieving its specific task in brain function.

ACTION POTENTIAL

An Increase in Na^+ and K^+ Conductance Generates Action Potentials

Hodgkin and Huxley not only recorded the action potential with an intracellular microelectrode (Fig. 5.1), but they also went on to perform a remarkable series of experiments that explained qualitatively and quantitatively the ionic mechanisms by which the action potential is generated (Hodgkin & Huxley, 1952a, 1952b). As mentioned earlier, these investigators found that during the action potential, the membrane potential of the cell rapidly overshoots 0 mV and approaches the equilibrium potential for Na^+. After generation of the action potential, the membrane potential repolarizes and becomes more negative than before, generating an after-hyperpolarization. These changes in membrane potential during generation of the action potential were associated with a large increase in conductance of the plasma membrane, but to what does the membrane become conductive in order to generate the action potential? The prevailing hypothesis was that there was a nonselective increase in conductance causing the negative resting potential to increase toward 0 mV. Since publication of the experiments of E. Overton in 1902, the action potential had been known to depend on the presence of extracellular Na^+. Reducing the concentration of Na^+ in the artificial seawater bathing the axon resulted in a marked reduction in the amplitude of the action potential. On the basis of these and other data, Hodgkin and Katz proposed that the action potential is generated through a rapid increase in the conductance of the membrane to Na^+ ions. A quantitative proof of this theory was lacking, however, because ionic currents could not be observed directly. Development of the voltage-clamp technique by Kenneth Cole at the Marine Biological Laboratory in Massachusetts resolved this problem and allowed quantitative measurement of the Na^+ and K^+ currents underlying the action potential (Cole, 1949; Box 5.3).

Hodgkin and Huxley used the voltage-clamp technique to investigate the mechanisms of generation of the action potential in the squid giant axon. Axons and neurons have a threshold for the initialization of an action potential of about −45 to −55 mV. Increasing the voltage from −60 to 0 mV produces a large, but transient, flow of positive charge into the cell (known as inward current). This transient inward current is followed by a sustained flow of positive charge out of the cell (the outward current). By

BOX 5.3

VOLTAGE-CLAMP TECHNIQUE

In the voltage-clamp technique, two independent electrodes are inserted into the squid giant axon: one for recording the voltage difference across the membrane and the other for intracellularly injecting the current (Fig. 5.6). These electrodes are then connected to a feedback circuit that compares the measured voltage across the membrane with the voltage desired by the experimenter. If these two values differ, then current is injected into the axon to compensate for this difference. This continuous feedback cycle, in which the voltage is measured and current is injected, effectively "clamps" the membrane at a particular voltage. If ionic channels were to open, then the resultant flow of ions into or out of the axon would be compensated for by the injection of positive or negative current into the axon through the current-injection electrode. The current injected through this electrode is necessarily equal to the current flowing through the ionic channels. It is this injected current that is measured by the experimenter. The benefits of the voltage-clamp technique are twofold. First, the current injected into the axon to keep the membrane potential "clamped" is necessarily equal to the current flowing through the ionic channels in the membrane, thereby giving a direct measurement of this current. Second, ionic currents are both voltage and time dependent; they become active at certain membrane potentials and do so at a particular rate. Keeping the voltage constant in the voltage clamp allows these two variables to be separated; the voltage dependence and the kinetics of the ionic currents flowing through the plasma membrane can be measured directly.

David A. McCormick

voltage clamping the cell and substituting different ions inside or outside the axon or both, Hodgkin, Huxley, and colleagues demonstrated that the transient inward current is carried by Na^+ ions flowing into the cell, and the sustained outward current is mediated by a sustained flux of K^+ ions moving out of the cell (Fig. 5.7) (Hille, 1977; Hodgkin & Huxley, 1952a, 1952b).

Na^+ and K^+ currents (I_{Na} and I_K, respectively) can be blocked, allowing each current to be examined in isolation (Fig. 5.7B). Tetrodotoxin (TTX), a powerful poison found in the puffer fish *Spheroides rubripes*, selectively blocks voltage-dependent Na^+ currents (the puffer fish remains a delicacy in Japan and must be prepared with the utmost care by the chef). Using TTX, one can selectively isolate I_K and examine its voltage dependence and time course (Fig. 5.7B).

Another compound, tetraethylammonium (TEA), is a useful pharmacological tool for selectively blocking I_K (Fig. 5.7B). The use of TEA to examine the voltage dependence and time course of the Na^+ current underlying action-potential generation (Fig. 5.7B) reveals some fundamental differences between Na^+ and K^+ currents. First, the inward Na^+ current activates, or "turns on," much more rapidly than the K^+ current (giving rise to the name "delayed rectifier" for this K^+ current). Second, the Na^+ current is transient; it inactivates, even if the membrane potential is maintained at 0 mV (Fig. 5.7A). In contrast, the outward K^+ current, once activated, remains "on" as long as the membrane potential is clamped to positive levels; that is, the K^+ current does not inactivate but is sustained. Remarkably, from one experiment, we see that the Na^+ current both activates and inactivates rapidly, whereas the K^+ current only activates slowly. These fundamental properties of the underlying Na^+ and K^+ channels allow the generation of action potentials.

Hodgkin and Huxley proposed that K^+ channels possess a voltage-sensitive "gate" that opens by depolarization and closes by the subsequent repolarization of the membrane potential. This process of "turning on" and "turning off" the K^+ current came to be known as activation and deactivation. The Na^+ current also exhibits voltage-dependent activation and deactivation (Fig. 5.7), but the Na^+ channels also become inactive despite maintained depolarization. Thus, the Na^+ current not only activates and deactivates, but it also exhibits a separate process known as inactivation, whereby the channels become blocked even though they are activated. Removal of this inactivation is achieved by removal of depolarization and is a process known as deinactivation. Thus, Na^+ channels possess two voltage-sensitive processes: activation–deactivation and inactivation–deinactivation.

The kinetics of these two properties of Na^+ channels are different; inactivation takes place at a slower rate than activation. The functional consequence of the two mechanisms is that Na^+ ions are allowed to flow across the membrane only when the current is activated but not inactivated. Accordingly, Na^+ ions do not flow at resting membrane potentials because the activation gate is closed (even though the inactivation gate is not). Upon depolarization, the activation gate opens, allowing Na^+ ions to flow into the cell. However, this depolarization also results in closure (at a slower rate) of the inactivation gate, which then blocks the flow of

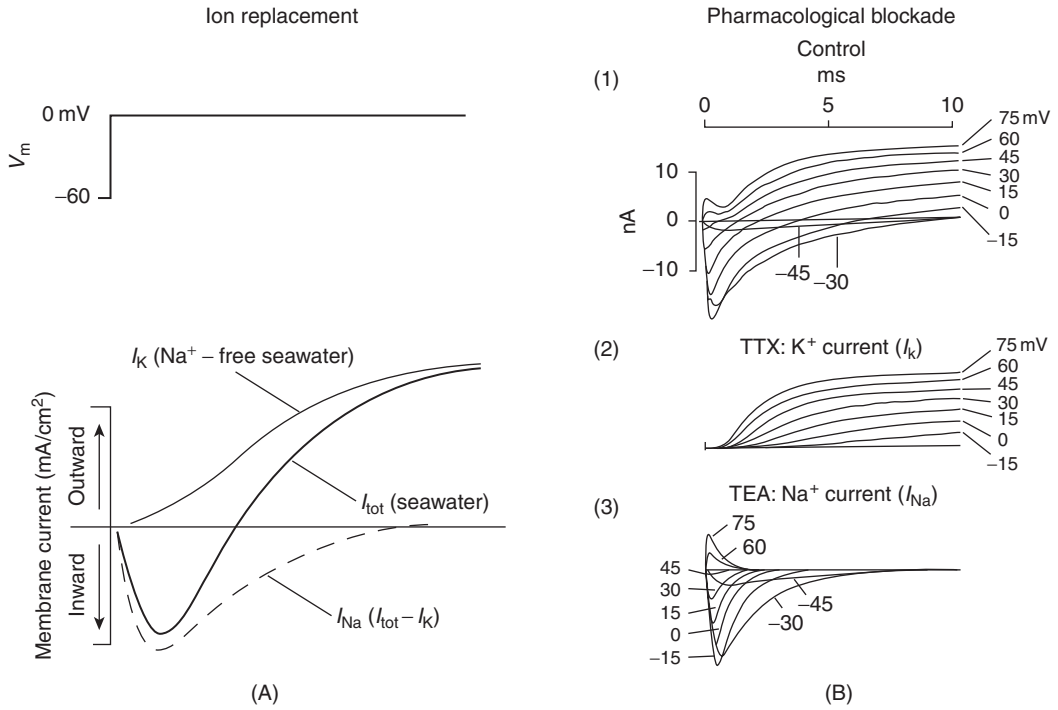


FIGURE 5.7 Voltage-clamp analysis reveals ionic currents underlying action potential generation. (A) Increasing the potential from −60 to 0 mV across the membrane of the squid giant axon activates an inward current followed by an outward current. If the Na^+ in seawater is replaced by choline (which does not pass through Na^+ channels), then increasing the membrane potential from −60 to 0 mV results in only the outward current, which corresponds to I_K. Subtracting I_K from the recording in normal seawater illustrates the amplitude–time course of the inward Na^+ current, I_{Na}. Note that I_K activates more slowly than I_{Na} and that I_{Na} inactivates with time. *From Hodgkin and Huxley (1952).* (B) These two ionic currents can also be isolated from one another through the use of pharmacological blockers. (1) Increasing the membrane potential from −45 to 75 mV in 15-mV steps reveals the amplitude–time course of inward Na^+ and outward K^+ currents. (2) After the block of I_{Na} with the poison tetrodotoxin (TTX), increasing the membrane potential to positive levels activates I_K only. (3) After the block of I_K with tetraethylammonium (TEA), increasing the membrane potential to positive levels activates I_{Na} only. *From Hille (1977).*

Na^+ ions. Upon repolarization of the membrane potential, the activation gate once again closes and the inactivation gate once again opens, preparing the axon for generation of the next action potential (Fig. 5.8)

Depolarization allows ionic current to flow by virtue of activation of the channel. The rush of Na^+ ions into the cell further depolarizes the membrane potential and more Na^+ channels become activated, forming a positive feedback loop that rapidly (within 100 μs or so) brings the membrane potential toward E_{Na}. However, the depolarization associated with generation of the action potential also inactivates Na^+ channels, and, as a larger and larger percentage of Na^+ channels become inactivated, the rush of Na^+ into the cell diminishes. This inactivation of Na^+ channels and the activation of K^+ channels result in the repolarization of the action potential. This repolarization deactivates the Na^+ channels. Then, the inactivation of the channel is slowly removed, and the channels are ready, once again, for the generation of another action potential (Fig. 5.8).

By measuring the voltage sensitivity and kinetics of these two processes, activation–deactivation and inactivation–deinactivation of the Na^+ current, as well as the activation–deactivation of the delayed rectifier K^+ current, Hodgkin and Huxley generated a series of mathematical equations that quantitatively described the generation of the action potential (calculation of the propagation of a single action potential required an entire week of cranking a mechanical calculator). According to these early experimental and computational neuroscientists, the action potential is generated as follows. Depolarization of the membrane potential increases the probability of Na^+ channels being in the activated but not yet inactivated, state. At a particular membrane potential, the resulting inflow of Na^+ ions tips the balance of the net ionic current from outward to inward (remember that depolarization will also increase K^+ and Cl^- currents by moving the membrane potential away from E_K and E_{Cl}). At this membrane potential, known as the action potential threshold (typically about −55 mV), the movement of Na^+ ions into the cell depolarizes the axon and opens more Na^+ channels, causing yet more depolarization of the membrane; repetition of this process yields a rapid, positive feedback loop that brings the axon close to E_{Na}. However, even as more and more Na^+ channels are becoming activated, some of these channels are also

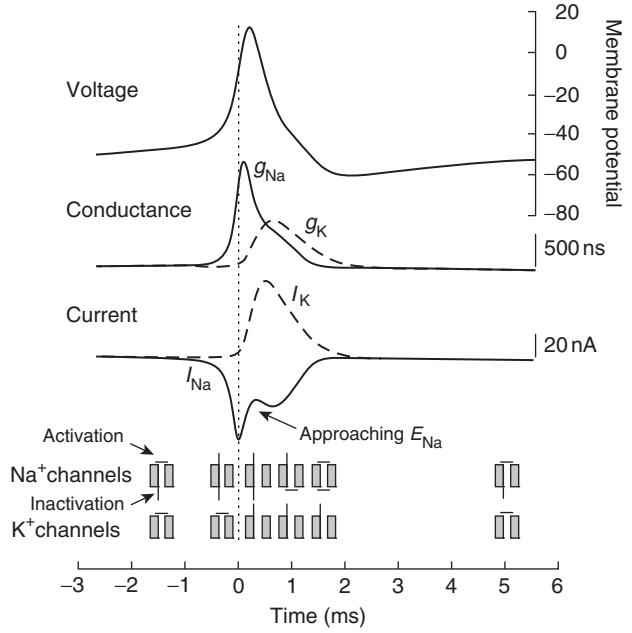


FIGURE 5.8 Generation of the action potential is associated with an increase in membrane Na^+ conductance and Na^+ current followed by an increase in K^+ conductance and K^+ current. Before action potential generation, Na^+ channels are neither activated nor inactivated (illustrated at the bottom of the figure). Activation of Na^+ channels allows Na^+ ions to enter the cell, depolarizing the membrane potential. This depolarization also activates K^+ channels. After activation and depolarization, the inactivation particle on the Na^+ channels closes and the membrane potential repolarizes. The persistence of the activation of K^+ channels (and other membrane properties) generates an after-hyperpolarization. During this period, the inactivation particle of the Na^+ channel is removed and the K^+ channels close.

inactivating and therefore no longer conducting Na^+ ions. In addition, the delayed rectifier K^+ channels are also opening, due to the depolarization of the membrane potential, and allowing positive charge to exit the cell. At some point, close to the peak of the action potential, the inward movement of Na^+ ions into the cell is exactly offset by the outward movement of K^+ ions out of the cell. After this point, the outward movement of K^+ ions dominates, and the membrane potential is repolarized, corresponding to the fall of the action potential. The persistence of the K^+ current for a few milliseconds following the action potential generates the after-hyperpolarization. During this after-hyperpolarization, which is lengthened by the membrane time constant, inactivation of the Na^+ channels is removed, preparing the axon for generation of the next action potential (Fig. 5.8).

The occurrence of an action potential is not associated with substantial changes in the intracellular or extracellular concentrations of Na^+ or K^+, as shown earlier for the generation of the resting membrane potential. For example, generation of a single action potential in a 25-μm-diameter hypothetical spherical cell should increase the intracellular concentration of Na^+ by only approximately 6 μM (from about 18 to 18.006 mM). Thus, the action potential is an electrical event generated by a change in the distribution of charge across the membrane and not by a marked change in the intracellular or extracellular concentration of Na^+ or K^+.

Action Potentials Typically Initiate in the Axon Initial Segment and Propagate Down the Axon and Backwards Through the Dendrites

Neurons have complex morphologies, including dendritic arbors, a cell body, and typically one axonal output that branches extensively. In many cells, all of these parts of the neuron are capable of independently generating action potentials. The activity of most neurons is dictated by barrages of synaptic potentials generated at each moment by a variable subset of the thousands of synapses impinging upon the cell's dendrites and soma. Where, then, is the action potential initiated? In most cells, each action potential is initiated in the initial portion of the axon, known as the axon initial segment (Coombs, Curtis, & Eccles, 1957; Popovic, Foust, McCormick, & Zecevic, 2011; Stuart, Spruston, Sakmann, & Hausser, 1997). The initial segment of the axon has the lowest threshold for action potential generation because it typically contains a moderately high density of Na^+ channels that have a more negative activation threshold and it is a small compartment that is easily depolarized by the in-rush of Na^+ ions. Once a spike is initiated (e.g., about 30–50 μm down the axon from the cell body in cortical pyramidal cells), this action potential then propagates orthodromically down the axon to the synaptic terminals, where it causes release of transmitter, as well as antidromically back through the cell body and into the cell dendrites, where it can modulate intracellular processes.

Refractory Periods Prevent "Reverberation"

The ability of depolarization to activate an action potential varies as a function of the time since the last generation of an action potential, due to the inactivation of Na^+ channels and the activation of K^+ channels. Immediately after the generation of an action potential, another action potential usually cannot be generated regardless of the amount of current injected into the axon. This period corresponds to the absolute refractory period and is largely mediated by the inactivation of Na^+ channels. The relative refractory period occurs during the action potential after-hyperpolarization and follows the absolute refractory period. The relative refractory period is characterized by a requirement for

the increased injection of ionic current into the cell to generate another action potential and results from persistence of the outward K^+ current. The practical implication of refractory periods is that action potentials are not allowed to "reverberate" between the soma and the axon terminals.

The Speed of Action Potential Propagation Is Affected by Myelination

Axons may be either myelinated or unmyelinated. Invertebrate axons or small vertebrate axons are typically unmyelinated, whereas larger vertebrate axons are often myelinated. As described in Chapter 3, sensory and motor axons of the peripheral nervous system are myelinated by specialized cells (Schwann cells) that form a spiral wrapping of multiple layers of myelin around the axon (Fig. 5.9). Several Schwann cells wrap around an axon along its length; between the ends of successive Schwann cells are small gaps (nodes of Ranvier). In the central nervous system, a single oligodendrocyte, a special type of glial cell, typically ensheaths several axonal processes.

In unmyelinated axons, the Na^+ and K^+ channels taking part in action potential generation are distributed along the axon, and the action potential propagates along the length of the axon through local depolarization of each neighboring patch of membrane, causing that patch of membrane to also generate an action potential (Fig. 5.9). Myelination of the axon reduces its membrane capacitance by moving the electrical charge differences between the inside and outside of the axon further apart and thus reducing their influence on each other. This has the effect of significantly increasing the passive length constant (λ) of the axon. In myelinated axons, however, the Na^+ channels are concentrated at the nodes of Ranvier. The generation of an action potential at each node results in depolarization of several adjacent nodes (owing to the long length constant of the myelinated axon) and subsequently generation of an action potential with an internode delay of only about 20 μs, referred to as saltatory conduction (from the Latin *saltare*, "to leap"). Interestingly, since action potentials are typically 1 msec or longer in duration, many adjacent nodes of Ranvier are in the process of generating action potentials at the same time (Popovic et al., 2011). Growing evidence indicates that, near the nodes of Ranvier and underneath the myelin covering, K^+ channels may play a role in determining the resting membrane potential and repolarization of the action potential. A cause of some neurological disorders, such as multiple sclerosis and Guillain-Barre syndrome, is the demyelination of axons, resulting in a block of conduction of the action potentials.

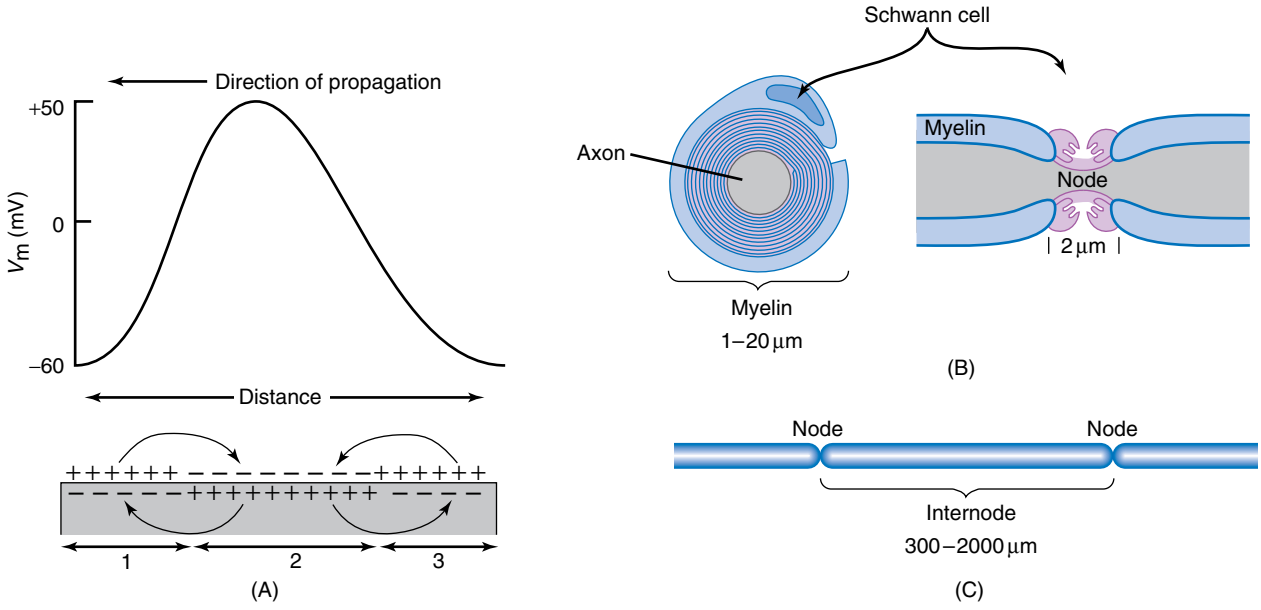


FIGURE 5.9 Propagation of the action potential in unmyelinated and myelinated axons. (A) Action potentials propagate in unmyelinated axons through the depolarization of adjacent regions of membrane. In the illustrated axon, region 2 is undergoing depolarization during the generation of the action potential, whereas region 3 has already generated the action potential and is now hyperpolarized. The action potential will propagate further by depolarizing region 1. (B) Vertebrate myelinated axons have a specialized Schwann cell that wraps around them in many spiral turns. The axon is exposed to the external medium at the nodes of Ranvier (Node). (C) Action potentials in myelinated fibers are regenerated at the nodes of Ranvier, where there is a high density of Na^+ channels. Action potentials are induced at each node through the depolarizing influence of the generation of an action potential at adjacent nodes, thereby increasing conduction velocity.

Ion Channels are Membrane-Spanning Proteins with Water-Filled Pores

The generation of ionic currents useful for the propagation of action potentials requires the movement of significant numbers of ions across the membrane in a relatively short time. The rate of ionic flow during the generation of an action potential is far too high to be achieved by an active transport mechanism and results instead from the opening of ion channels. Although the existence of ionic channels in the membrane has been postulated for decades, their properties and structure have only recently become known in detail. The powerful combination of electrophysiological and molecular techniques has enhanced our knowledge of the structure–function relations of ionic channels greatly (Box 5.4).

Various neural toxins were particularly useful in the initial isolation of ionic channels. For example, three subunits (α, $\beta 1$, $\beta 2$) of the voltage-dependent Na^+ channel were isolated with the use of a derivative of a scorpion toxin. The α subunit of the Na^+ channel is a large glycoprotein with a molecular mass of 270 kDa, whereas the $\beta 1$ and $\beta 2$ subunits are smaller polypeptides of molecular masses 39 and 37 kDa, respectively (Fig. 5.10). The α subunit, of which there are at least nine different isoforms, is the building block of the water-filled pore of the ionic channel, whereas the β subunits have some other role, such as in the regulation or structure of the native channel (Catterall, 2000a).

The α subunit of the Na^+ channel contains four internal repetitions (Fig. 5.10B). Hydrophobicity analysis of these four components reveals that each contains six

BOX 5.4

ION CHANNELS AND DISEASE

Cells cannot survive without functional ion channels. It is therefore not surprising that an ever-increasing number of diseases have been found to be associated with defective ion channel function. There are a number of different mechanisms by which this may occur.

1. Mutations in the coding region of ion channel genes may lead to gain or loss of channel function, either of which may have deleterious consequences. For example, mutations producing enhanced activity of the epithelial Na^+ channel are responsible for Liddle's syndrome, an inherited form of hypertension, whereas other mutations in the same protein that cause reduced channel activity give rise to hypotension. The most common inherited disease in Caucasians is also an ion channel mutation. This disease is cystic fibrosis (CF), which results from mutations in the epithelial chloride channel, known as CFTR. The most common mutation, deletion of a phenylalanine at position 508, results in defective processing of the protein and prevents it from reaching the surface membrane. CFTR regulates chloride fluxes across epithelial cell membranes, and this loss of CFTR activity leads to reduced fluid secretion in the lung, resulting in potentially fatal lung infections.
2. Mutations in the promoter region of the gene may cause under- or overexpression of a given ion channel.
3. Other diseases result from defective regulation of channel activity by cellular constituents or extracellular ligands. This defective regulation may be caused by mutations in the genes encoding the regulatory molecules themselves or defects in the pathways leading to their production. Some forms of maturity-onset diabetes of the young (MODY) may be attributed to such a mechanism. ATP-sensitive potassium (K-ATP) channels play a key role in the glucose-induced insulin secretion from pancreatic β cells, and their defective regulation is responsible for one form of MODY.
4. Autoantibodies to channel proteins may cause disease by downregulating channel function—often by causing internalization of the channel protein itself. Well-known examples are myasthenia gravis, which results from antibodies to skeletal muscle acetylcholine channels, and Lambert–Eaton myasthenic syndrome, in which patients produce antibodies against presynaptic Ca^{2+} channels.
5. Finally, a number of ion channels are secreted by cells as toxic agents. They insert into the membrane of the target cell and form large nonselective pores, leading to cell lysis and death. The hemolytic toxin produced by the bacterium *Staphylococcus aureus* and the toxin secreted by the protozoan *Entamoeba histolytica*, which causes amebic dysentery, are examples.

Natural mutations in ion channels have been invaluable for studying the relationship between channel structure and function. In many cases, genetic analysis of a disease has led to the cloning of the relevant ion channel. The first K^+ channel to be identified (Shaker), for example, came from the cloning of the gene that caused *Drosophila* to shake when exposed to ether. Likewise, the gene encoding the primary subunit of a cardiac potassium channel (KCNQ1) was identified by positional cloning in families carrying mutations that caused a cardiac disorder known

BOX 5.4 (*cont'd*)

as long QT syndrome (see later). Conversely, the large number of studies on the relationship between Na^+ channel structure and function has greatly assisted our understanding of how mutations in Na^+ channels produce their clinical phenotypes.

Many diseases are genetically heterogeneous, and the same clinical phenotype may be caused by mutations in different genes. Long QT syndrome is a relatively rare inherited cardiac disorder that causes abrupt loss of consciousness, seizures, and sudden death from ventricular arrhythmia in young people. Mutations in five different genes, two types of cardiac muscle K^+ channels (HERG, KCNQ1, KCNE1, KCNE2) and the cardiac muscle sodium channel (SCN1A), give rise to long QT syndrome. The disorder is characterized by a long QT interval in the electrocardiogram, which reflects the delayed repolarization of the cardiac action potential. As might therefore be expected, mutations in the cardiac Na^+ channel gene that cause long QT syndrome enhance the Na^+ current (by reducing Na^+ channel inactivation), whereas those in potassium channel genes cause loss of function and reduce the K^+ current.

Mutations in many different types of ion channels have been shown to cause human diseases. In addition to the examples listed earlier, mutations in water channels cause nephrogenic diabetes insipidus; mutations in gap junction channels cause Charcot–Marie–Tooth disease (a form of peripheral neuropathy) and hereditary deafness; mutations in the skeletal muscle Na^+ channel cause a range of disorders known as periodic paralyses; mutations in intracellular Ca^{2+}-release channels cause malignant hyperthermia (a disease in which inhalation anesthetics trigger a potentially fatal rise in body temperature); and mutations in neuronal voltage-gated Ca^{2+} channels cause migraine and episodic ataxia. The list increases daily. As is the case with all single gene disorders, the frequency of these diseases in the general population is very low. However, the insight they have provided into the relationship between ion channel structure and function, and into the physiological role of the different ion channels, has been invaluable. As William Harvey said in 1657, "Nor is there any better way to advance the proper practice of medicine than to give our minds to the discovery of the usual form of nature, by careful investigation of the rarer forms of disease."

Frances M. Ashcroft

hydrophobic domains that may span the membrane as an α-helix. Of these six membrane-spanning components, the fourth (S4) has been proposed to be critical to the voltage sensitivity of the Na^+ channels. Voltage-sensitive gating of Na^+ channels is accomplished by the redistribution of ionic charge ("gating charge") in the channel. Positive charges in the S4 region may act as voltage sensors such that an increase in the positivity of the inside of the cell results in a conformational change of the ionic channel. In support of this hypothesis, site-directed mutagenesis of the S4 region of the Na^+ channel to reduce the positive charge of this portion of the pore also reduces the voltage sensitivity of activation of the ionic channel.

The mechanisms of inactivation of ionic channels have been analyzed with a combination of molecular and electrophysiological techniques. The most convincing hypothesis is that inactivation is achieved by a block of the inner mouth of the aqueous pore. Ionic channels are inactivated without detectable movement of ionic current through the membrane; thus inactivation is probably not directly gated by changes in the membrane potential alone. Rather, inactivation is triggered or facilitated as a secondary consequence of activation. Site-directed mutagenesis or the use of antibodies has shown that the part of the molecule between regions III and IV may be allowed to move to block the cytoplasmic side of the ionic pore after the conformational change associated with activation.

Neurons of the Central Nervous System Exhibit a Wide Variety of Electrophysiological Properties

The first intracellular recordings of action potentials in mammalian neurons by Sir John Eccles and colleagues revealed a remarkable similarity to those of the squid giant axon and gave rise to the assumption that the electrophysiology of neurons in the CNS was really rather simple: when synaptic potentials brought the membrane potential positive to action potential threshold, action potentials were produced through an increase in Na^+ conductance followed by an increase in K^+ conductance, as in the squid giant axon. The assumption, therefore, was that the complicated patterns of activity generated by the brain during the resting, sleeping, or active states were brought about as an interaction of the very large numbers of neurons present in the mammalian CNS. However, intracellular recordings of invertebrate neurons revealed that different cell types exhibit a wide variety of different electrophysiological behaviors, indicating that neurons may be significantly more complicated than the squid giant axon.

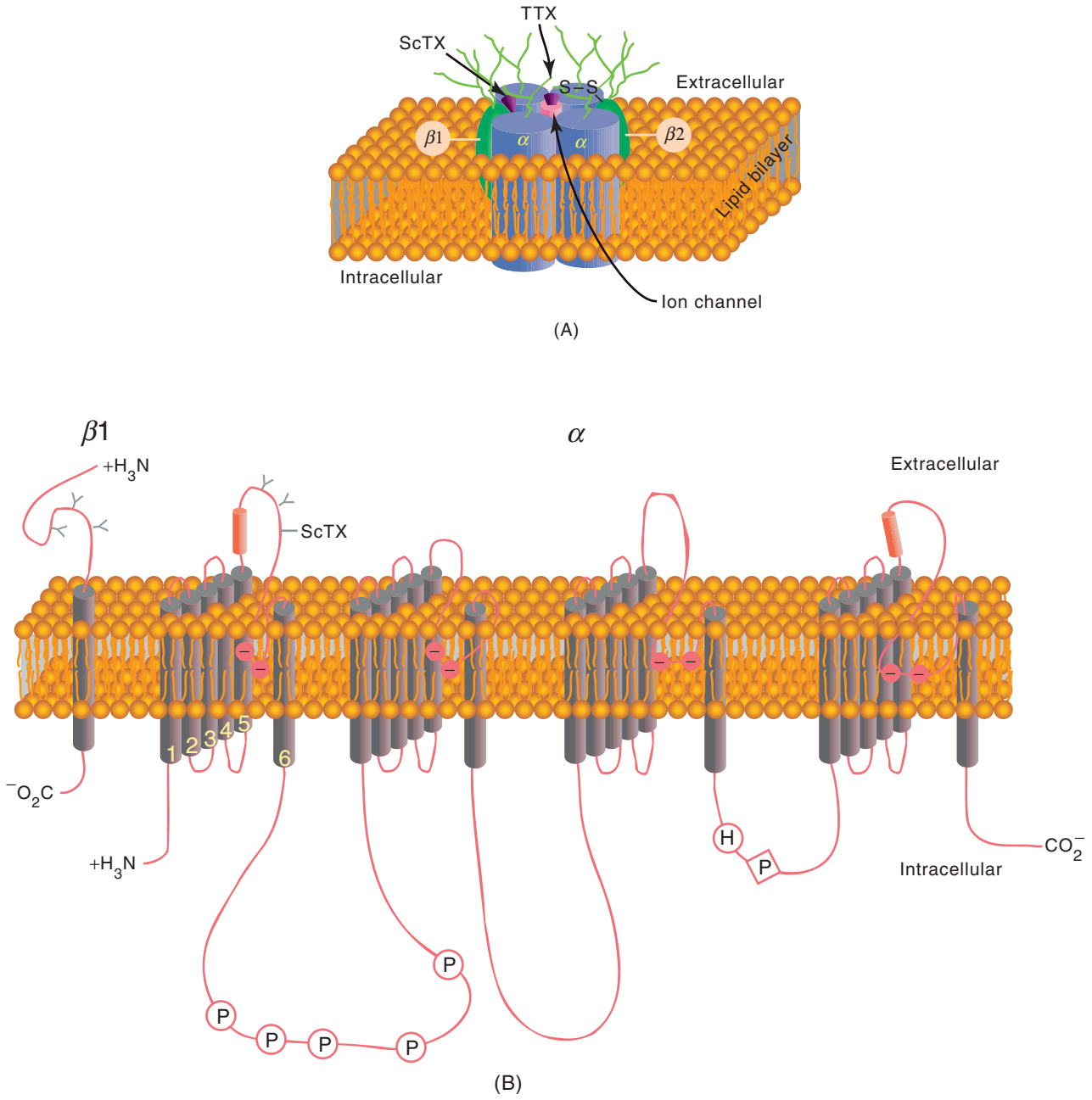


FIGURE 5.10 Structure of the sodium channel. (A) Cross section of a hypothetical sodium channel consisting of a single transmembrane α subunit in association with a $\beta 1$ subunit and a $\beta 2$ subunit. The α subunit has receptor sites for α-scorpion toxins (ScTX) and tetrodotoxin (TTX). (B) Primary structures of α and $\beta 1$ subunits of sodium channel illustrated as transmembrane-folding diagrams. Cylinders represent probable transmembrane α-helices.

Elucidation of the basic electrophysiology and synaptic physiology of different types of neurons and neuronal pathways within the mammalian CNS was facilitated by the *in vitro* slice technique, in which thin (~0.5 mm) slices of brain can be maintained for several hours. Intracellular recordings from identified cells revealed that neurons of the mammalian nervous system, such as those of invertebrate networks, can generate complex patterns of action potentials entirely through intrinsic ionic mechanisms and without synaptic interaction with other cell types. For example, Rodolfo Llinás and colleagues discovered that Purkinje cells of the cerebellum can generate high-frequency trains (>200 Hz) of Na^+- and K^+-mediated action potentials interrupted by Ca^{2+} spikes in the dendrites, whereas a major afferent to these neurons, the inferior olivary cell, can generate rhythmic sequences of

broad action potentials only at low frequencies (<15 Hz) through an interaction between various Ca^{2+}, Na^{+}, and K^{+} conductances (Fig. 5.11). These *in vitro* recordings confirmed a major finding obtained with earlier intracellular recordings *in vivo*: each morphologically distinct class of neuron in the brain exhibits distinct electrophysiological features. Just as cortical pyramidal cells are morphologically distinct from cerebellar

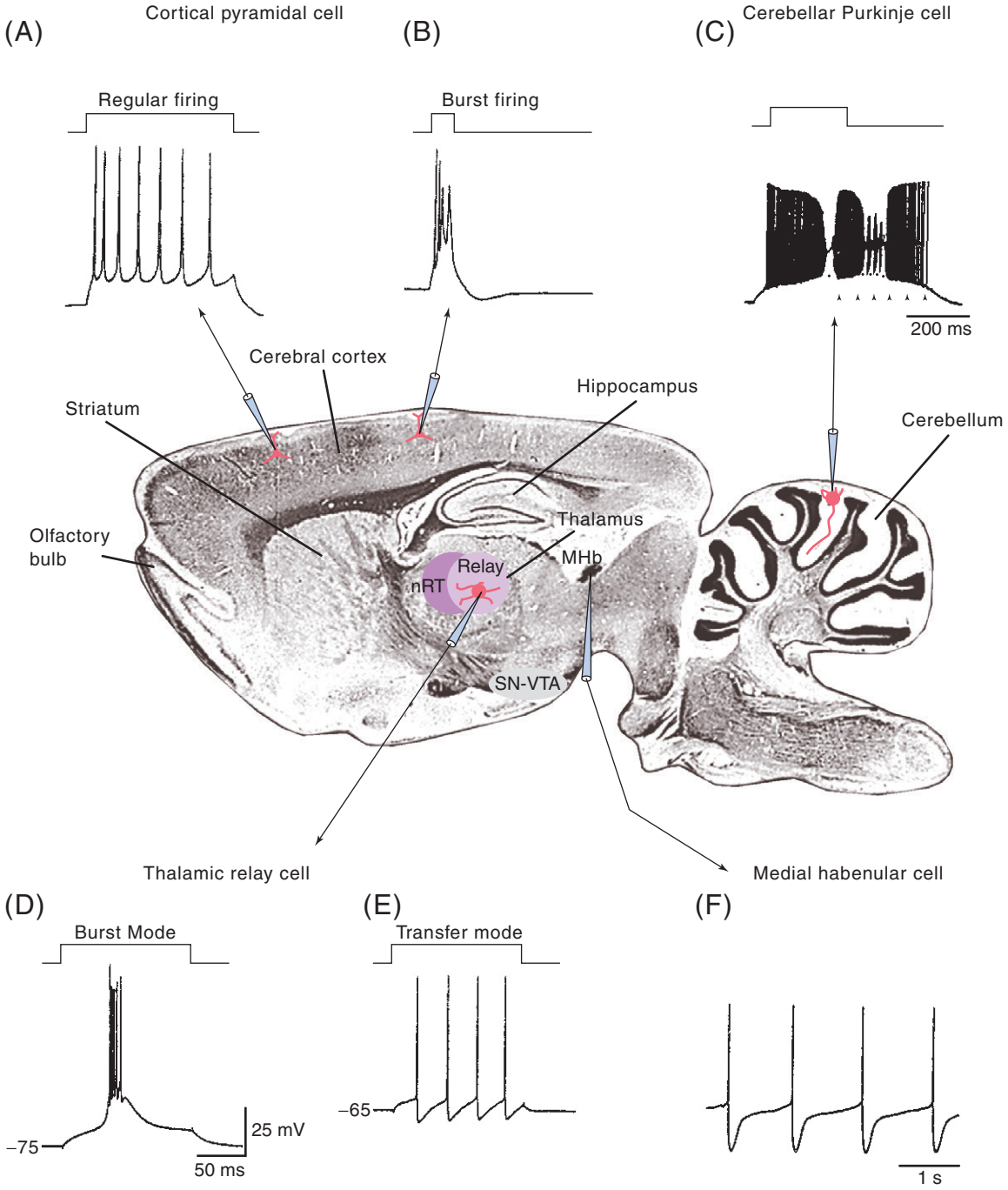


FIGURE 5.11 Neurons in the mammalian brain exhibit widely varying electrophysiological properties. (A) Intracellular injection of a depolarizing current pulse in a cortical pyramidal cell results in a train of action potentials that slow down in frequency. This pattern of activity is known as "regular firing." (B) Some cortical cells generated bursts of three or more action potentials, even when depolarized only for a short period of time. (C) Cerebellar Purkinje cells generate high-frequency trains of action potentials in their cell bodies that are disrupted by the generation of Ca^{2+} spikes in their dendrites. These cells can also generate "plateau potentials" from the persistent activation of Na^{+} conductances (arrowheads). Thalamic relay cells may generate action potentials either as bursts (D) or as tonic trains of action potentials (E) due to the presence of a large low-threshold Ca^{2+} current. (F) Medial habenular cells generate action potentials at a steady and slow rate in a "pacemaker" fashion.

Purkinje cells, which are distinct from thalamic relay cells, the electrophysiological properties of each of these different cell types are also markedly distinct (Llinás, 1988).

Although no uniform classification scheme has been formulated in which all the different types of neurons of the brain can be classified, a few characteristic patterns of activity seem to recur. The first general class of action potential generation is characterized by those cells that generate trains of action potentials one spike at a time. The more prolonged the depolarization of these cells, the more prolonged their discharge. The more intensely these cells are depolarized, the higher the frequency of action potential generation. This type of relatively linear behavior is typical for brainstem and spinal cord motor neurons functioning in muscle contraction. A modification of this basic pattern of "regular firing" is characterized by the generation of trains of action potentials that exhibit a marked tendency to slow down in frequency with time, a process known as spike frequency adaptation. Examples of cells that discharge in this manner are cortical and hippocampal pyramidal cells.

In addition to these regular firing cells, many neurons in the central nervous system exhibit the intrinsic propensity to generate rhythmic bursts of action potentials (Fig. 5.11). Examples of such neurons are thalamic relay neurons, inferior olivary neurons, and some types of cortical and hippocampal pyramidal cells. In these cells, clusters of action potentials can occur together when the membrane is brought above the firing threshold. These clusters of action potentials are typically generated through the activation of specialized Ca^{2+} currents that, through their slower kinetics, allow the membrane potential to be depolarized for a sufficient period to result in the generation of a burst of regular, Na^+- and K^+-dependent action potentials (discussed in the next section).

Yet another general category of neurons in the brain comprises cells that generate relatively short duration (<1 ms) action potentials and can discharge at relatively high frequencies (>300 Hz). Such electrophysiological properties are often found in neurons that release the inhibitory amino acid γ-aminobutyric acid (Fig. 5.11), including Purkinje cells of the cerebellum and some types of interneurons in the cerebral cortex, thalamus, and hippocampus. Finally, the last general category of neurons consists of those that spontaneously generate action potentials at relatively slow frequencies (e.g., 1–10 Hz). This type of electrophysiological behavior is often associated with neurons that release neuromodulatory transmitters, such as acetylcholine, norepinephrine, serotonin, and histamine. Neurons that release these neuromodulatory substances often innervate wide regions of the brain and appear to set the "state" of the different neural networks of the CNS in a manner similar to the modulation of the different organs of the body by the sympathetic and parasympathetic nervous systems.

Each of these unique intrinsic patterns of activity in the nervous system is due to the presence of a distinct mixture and distribution of different ionic currents in the cells. As in classical studies of the squid giant axon, these different ionic currents have been characterized, at least in part, with voltage-clamp and pharmacological techniques, and the basic electrophysiological properties have been replicated with computational simulations (Figs. 5.8 and 5.13).

Neurons Have Multiple Active Conductances

The search for the electrophysiological basis of the varying intrinsic properties of different types of neurons of vertebrates and invertebrates revealed a wide variety of ionic currents. Each type of ionic current is characterized by several features: (1) the type of ions conducted by the underlying ionic channels (e.g., Na^+, K^+, Ca^{2+}, Cl^-, or mixed cations), (2) their voltage and time dependence, and (3) their sensitivity to second messengers. In vertebrate neurons, two distinct Na^+ currents have been identified and six distinct Ca^{2+} currents and more than seven distinct K^+ currents are known (Fig. 5.12). This is a minimal number, since these currents are formed from a much greater pool of channel subunits. The following sections briefly review these classes of ionic currents and their ionic channels, relating them to the different patterns of behavior mentioned earlier for neurons in the mammalian CNS.

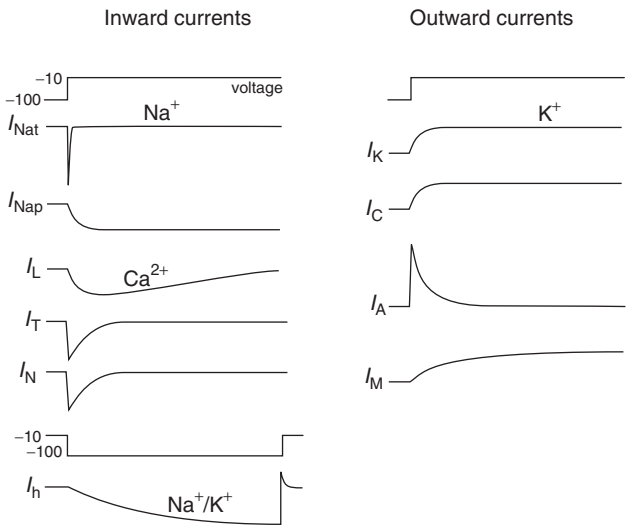


FIGURE 5.12 Voltage dependence and kinetics of different ionic currents in the mammalian brain. Depolarization of the membrane potential from –100 to –10 mV results in the activation of currents entering or leaving neurons.

Na^+ Currents are Both Transient and Persistent

Depolarization of many different types of vertebrate neurons results not only in the activation of the rapidly activating and inactivating Na^+ current (I_{Nat}) underlying action potential generation, but also in the rapid activation of a Na^+ current that does not inactivate and is therefore known as the "persistent" Na^+ current (I_{Nap}). The threshold for activation of the persistent Na^+ current is typically about −65 mV—that is, below the threshold for the generation of action potentials. This property gives this current the interesting ability to enhance or facilitate the response of the neuron to depolarizing, yet subthreshold, inputs. For example, synaptic events that depolarize the cell will activate I_{Nap}, resulting in an extra influx of positive charge and therefore a larger depolarization than otherwise would occur. Likewise, hyper-polarizations may result in deactivation of I_{Nap}, again resulting in larger hyperpolarizations than would otherwise occur. In this manner, the persistent Na^+ current may play an important regulatory function in the control of the functional responsiveness of the neuron to synaptic inputs and may contribute to the dynamic coupling of the dendrites to the soma.

Persistent activation of I_{Nap} may also contribute to another electrophysiological feature of neurons: the generation of plateau potentials. A plateau potential refers to the ability of many different types of neurons to generate, through intrinsic ionic mechanisms, a prolonged (from tens of milliseconds to seconds) depolarization and action potential discharge in response to a short-lasting depolarization (Fig. 5.11C). One can wonder whether such plateau potentials contribute to persistent firing in neurons during the performance of working memory tasks, where groups of neurons have to retain information for brief (seconds) periods of time.

K^+ Currents Vary in Their Voltage Sensitivity and Kinetics

Potassium currents that contribute to the electrophysiological properties of neurons are numerous and exhibit a wide range of voltage-dependent and kinetic properties (reviewed in Coetzee et al., 1999). Perhaps the simplest K^+ current is that characterized by Hodgkin and Huxley: this K^+ current, I_K, activates rapidly on depolarization and does not inactivate (Fig. 5.12). Other K^+ currents activate with depolarization but also inactivate with time. For example, the rapid activation and inactivation of I_A give this current a transient appearance (Fig. 5.12), and I_A is believed to be important in controlling the rate of action potential generation, particularly at low frequencies (Fig. 5.13). Like the Na^+ channel, I_A channels are inactivated by the plugging of the inner mouth of the pore through movement of an inactivation particle.

Another broad class of K^+ channels consists of those that are sensitive to changes in the intracellular concentration of Ca^{2+}. These K^+ currents are collectively referred to as I_{KCa} (Fig. 5.12). Still other K^+ channels are not only activated by voltage but are also modulated by the activation of various modulatory neurotransmitter receptors (e.g., I_M; Fig. 5.12). Between these classic examples of K^+ currents are a variety of other types that have not been fully characterized, including K^+ currents that vary from one another in their voltage sensitivity, kinetics, and response to various second messengers.

Molecular biological studies of voltage-sensitive K^+ channels, first done in Drosophila and later in mammals, have revealed the presence of a large number of genes that generate K^+ channels. The voltage-gated K^+ channel subunits are contained within nine distinct subfamilies: Kv1-9 (reviewed in Coetzee et al., 1999). These genes generate a wide variety of different K^+ channels, not only due to the large number of genes involved, but also due to alternative RNA splicing, gene duplication, and other mechanisms. Functional expression of different K^+ channels reveals remarkable variation in the rate of inactivation, such that some are rapidly inactivating (A-current like), whereas others inactivate more slowly and, finally, some K^+ channels do not inactivate, such as I_K. One of the largest subfamilies of K^+ channels are those that give rise to the resting membrane potential, so-called "leak channels." Interestingly, these channels appear to be opened by gaseous anesthetics, indicating that hyperpolarization of central neurons is a major component of general anesthesia. It is now clear that each type of neuron in the nervous system contains a unique set of functional voltage-sensitive K^+ channels, selected, modified, and placed in particular spatial locations in the cell in a manner that facilitates the unique role of that cell type in neuronal processing.

An additional current that also regulates the responsiveness of neurons to depolarizing inputs is the voltage-sensitive K^+ current known as the M current (Figs. 5.12 and 5.13D) (Brown & Passmore, 2009). By investigating the ionic mechanisms by which the release of acetylcholine from preganglionic neurons in the brain results in prolonged changes in the excitability of neurons of the sympathetic ganglia, Brown and Adams (1980) discovered a unique K^+ current that slowly (over tens of milliseconds) turns on with depolarization of the neuron (Fig. 5.13D). The slow activation of this K^+ current results in a decrease in the responsiveness of the cell to depolarization and therefore regulates how the cell responds to excitation. This K^+ current, like I_{AHP}, is reduced by the activation of a wide variety of receptors, including muscarinic receptors, for which it is named. Reduction of I_M results in a marked increase in responsiveness of the affected cell to depolarizing inputs and

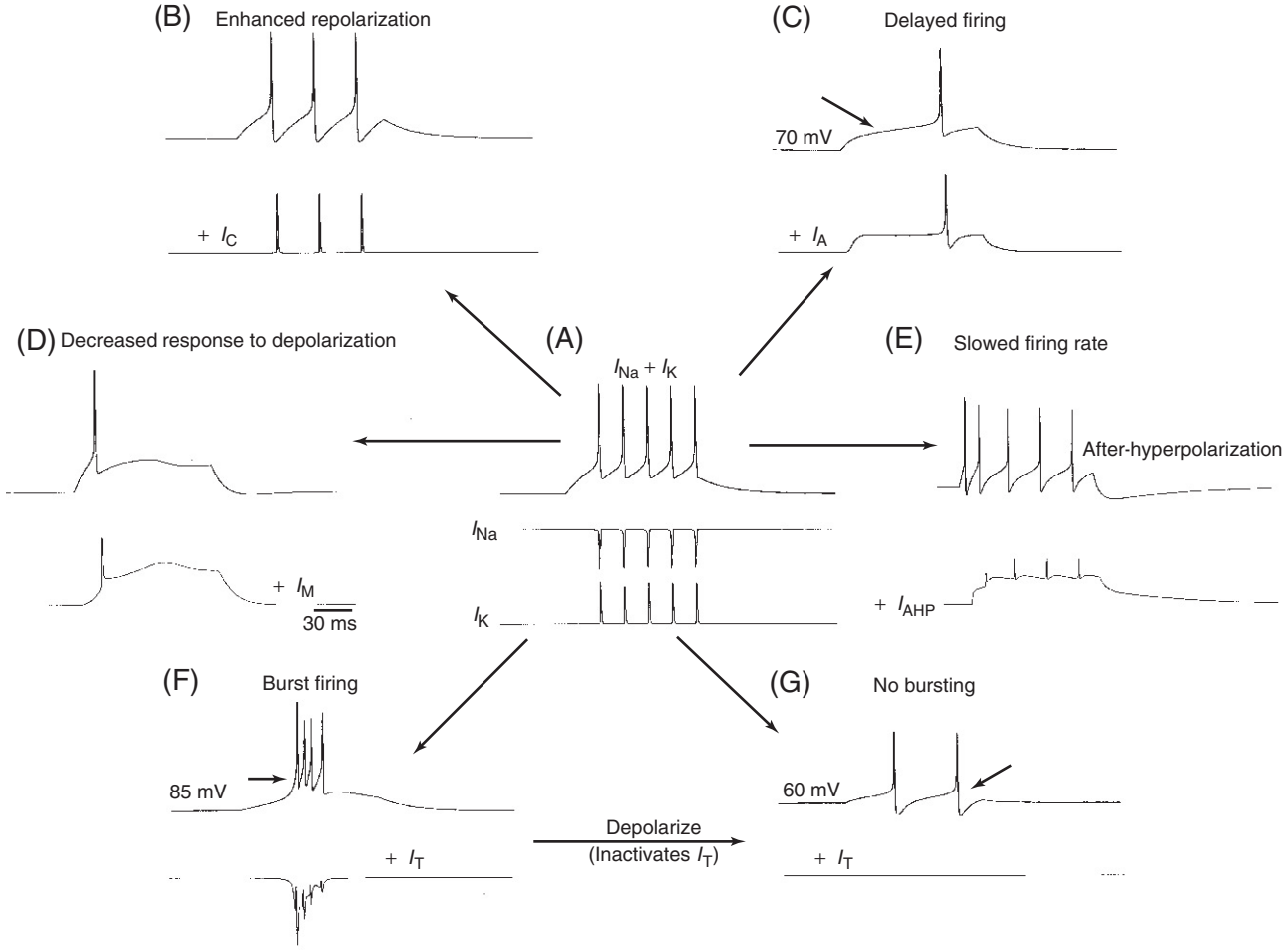


FIGURE 5.13 Simulation of the effects of the addition of various ionic currents to the pattern of activity generated by neurons in the mammalian CNS. (A) The repetitive impulse response of the classical Hodgkin–Huxley model (voltage recordings above, current traces below). With only I_{Na} and I_K, the neuron generates a train of five action potentials in response to depolarization. Addition of I_C (B) enhances action potential repolarization. Addition of I_A (C) delays the onset of action potential generation. Addition of I_M (D) decreases the ability of the cell to generate a train of action potentials. Addition of I_{AHP} (E) slows the firing rate and generates a slow after-hyperpolarization. Finally, addition of the transient Ca^{2+} current I_T results in two states of action potential firing: (F) burst firing at −85 mV and (G) tonic firing at −60 mV. *From Huguenard and McCormick (1994).*

again may contribute to the mechanisms by which neuromodulatory systems control the state of activity in cortical and hippocampal networks (reviewed in McCormick, 1992; Nicoll, 1988).

Ca^{2+} Currents Control Electrophysiological Properties and Ca^{2+}-Dependent Second-Messenger Systems

Ionic channels that conduct Ca^{2+} are present in all neurons. These channels are special in that they serve two important functions. First, Ca^{2+} channels are present throughout the different parts of the neuron (dendrites, soma, synaptic terminals) and contribute greatly to the electrophysiological properties of these processes. Second, Ca^{2+} channels are unique in that Ca^{2+} is an important second messenger in neurons, and entry of Ca^{2+} into the cell can affect numerous physiological functions, including neurotransmitter release, synaptic plasticity, neurite outgrowth during development, and even gene expression. On the bases of their voltage sensitivity, their kinetics of activation and inactivation, and their ability to be blocked by various pharmacological agents, Ca^{2+} currents can be separated into at least six separate categories, three of which are I_T ("transient"), I_L ("long lasting"), and I_N ("neither"), illustrated in Figure 5.12A. A fourth, I_P, is found in Purkinje cells of the cerebellum, as well as in many different cell types of the CNS. These Ca^{2+} channels are formed from at least 10 different α subunits as well as a variety of β and γ subunits, indicating that an even greater number of Ca^{2+} currents are yet to be characterized.

Neurons Possess Multiple Subtypes of High-Threshold Ca^{2+} Currents

High-voltage-activated Ca^{2+} channels are activated at membrane potentials positive to approximately −40 mV and include the currents I_L, I_N, and I_P. L-type calcium currents exhibit a high threshold for activation (about −10 mV) and give rise to rather persistent, or long-lasting, ionic currents (Fig. 5.12A). Dihydropyridines, Ca^{2+} channel antagonists, are clinically useful for their effects on the heart and vascular smooth muscle (e.g., for the treatment of arrhythmias, angina, and migraine headaches) and selectively block L-type Ca^{2+} channels. In contrast with I_L, I_N is not blocked by dihydropyridines: it is blocked selectively by a toxin found in Pacific cone shells (ω-conotoxin-GVIA). N-type Ca^{2+} channels have a threshold for activation of about −20 mV, inactivate with maintained depolarization, and are modulated by a variety of neurotransmitters. In some cell types, I_N has a role in the Ca^{2+}-dependent release of neurotransmitters at presynaptic terminals. The P-type calcium channel is distinct from N and L types in that it is not blocked by either dihydropyridines or ω-conotoxin-GVIA but is blocked by a toxin (ω-agatoxin-IVA) present in the venom of the Funnel web spider. This type of calcium channel activates at relatively high thresholds and does not inactivate. Prevalent in Purkinje cells, as well as other cell types, as mentioned earlier, the P-type Ca^{2+} channel participates in the generation of dendritic Ca^{2+} spikes, which can strongly modulate the firing pattern of the neuron in which it resides (Fig. 5.11C; Llinás, 1988).

Collectively, high-threshold-activated Ca^{2+} channels contribute to the generation of action potentials in mammalian neurons. The activation of Ca^{2+} currents adds somewhat to the depolarizing part of the action potential, but, more importantly, these channels allow Ca^{2+} to enter the cell, which has the secondary consequence of activation of various Ca^{2+}-activated K^+ currents and protein kinases (see Chapter 8). As mentioned earlier, activation of these K^+ currents modifies the pattern of action potentials generated in the cell (Figs. 5.11 and 5.13).

High-threshold Ca^{2+} channels are similar to the Na^+ channel in that they are composed of a central α1 subunit that forms the aqueous pore and several regulatory or auxiliary subunits. As in the Na^+ channel, the primary structure of the α1 subunit of the Ca^{2+} channel consists of four homologous domains (I–IV), each containing six regions (S1–S6) that may generate transmembrane α-helices. Genes for at least 10 different Ca^{2+} channel α subunits have been cloned and are separated into three subfamilies (Cav1, Cav2, and Cav3). The properties of the products of these genes indicate that I_L is likely to correspond to the Cav1 subfamily, whereas I_N corresponds to Cav2.2 and I_T is formed from the Cav3 subfamily (see Catterall, 2000b).

Low-Threshold Ca^{2+} Currents Generate Bursts of Action Potentials

Low-threshold Ca^{2+} currents (Fig. 5.12A) often take part in the generation of rhythmic bursts of action potentials (Figs. 5.11 and 5.13). The low-threshold Ca^{2+} current is characterized by a threshold for activation of about −65 mV, which is below the threshold for generation of typical Na^+–K^+-dependent action potentials (−55 mV). This current inactivates with maintained depolarization. Because of these properties, the role of low-threshold Ca^{2+} currents differs markedly from that of the high-threshold Ca^{2+} currents. Through activation and inactivation of the low-threshold Ca^{2+} current, neurons can generate slow (about 100 ms) Ca^{2+} spikes, which can result, due to their prolonged duration, in generation of a high-frequency "burst" of short-duration Na^+–K^+ action potentials (Fig. 5.11).

In the mammalian brain, this pattern is especially well exemplified by the activity of thalamic relay neurons; in the visual system, these neurons receive direct input from the retina and transmit this information to the visual cortex. During periods of slow wave sleep, the membrane potential of these relay neurons is relatively hyperpolarized, resulting in the removal of inactivation (deinactivation) of the low-threshold Ca^{2+} current. This deinactivation allows these cells to spontaneously generate low-threshold Ca^{2+} spikes and bursts of from two to five action potentials (Fig. 5.14). The large number of thalamic relay cells bursting during sleep in part gives rise to the spontaneous synchronized activity that early investigators were so surprised to find during recordings from the brains of sleeping animals. It has even proved possible to maintain one of the sleep-related brain rhythms (spindle waves) intact in slices of thalamic tissue maintained *in vitro*, due to the generation of this rhythm by the interaction of a local network of thalamic cells and their electrophysiological properties.

The transition to waking or the period of sleep when dreams are prevalent (rapid eye movement sleep) is associated with a maintained depolarization of thalamic relay cells to membrane potentials ranging from about −60 to −55 mV. The low-threshold Ca^{2+} current is inactivated and therefore the burst discharges are abolished. In this way, the properties of a single ionic current (I_T) help explain in part the remarkable changes in brain activity taking place in the transition from sleep to waking (Fig. 5.14).

Hyperpolarization-Activated Ionic Currents Are Involved in Rhythmic Activity

In most types of neurons, hyperpolarization negative to approximately −60 mV activates an ionic current, known as I_h, that conducts both Na^+ and K^+ ions (Fig. 5.12). This

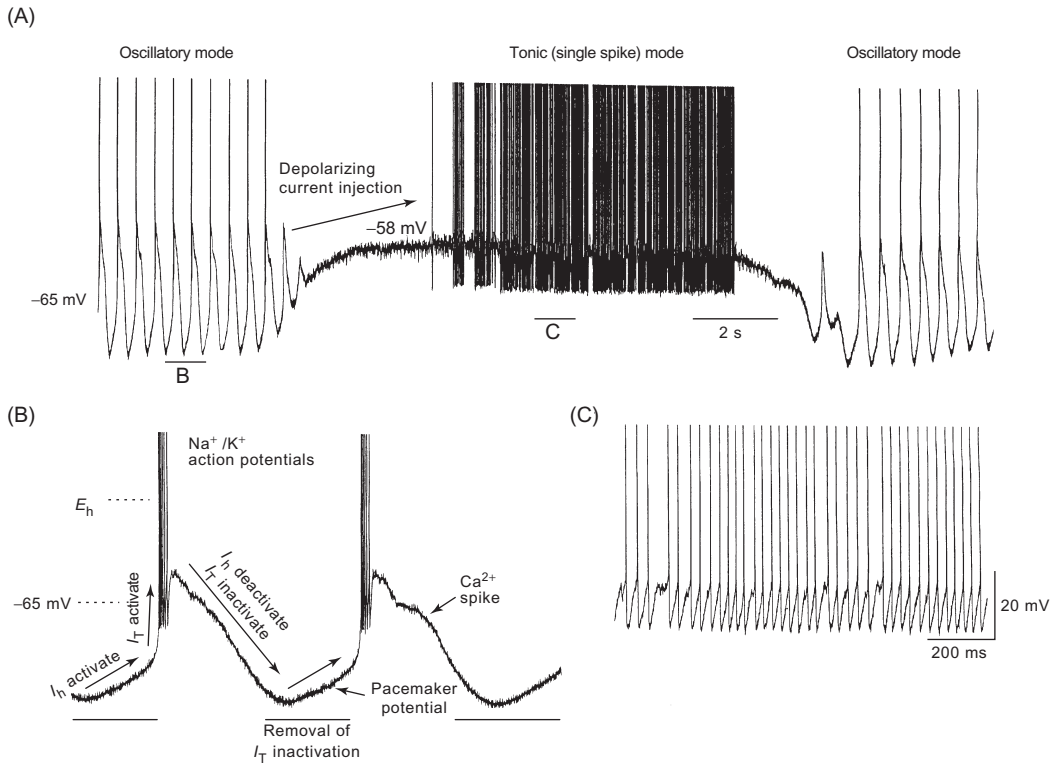


FIGURE 5.14 Two different patterns of activity generated in the same neuron, depending on membrane potential. (A) The thalamic neuron spontaneously generates rhythmic bursts of action potentials due to the interaction of the Ca^{2+} current I_T and the inward "pacemaker" current I_h. Depolarization of the neuron changes the firing mode from rhythmic burst firing to tonic action potential generation in which spikes are generated one at a time. Removal of this depolarization reinstates rhythmic burst firing. This transition from rhythmic burst firing to tonic activity is similar to that which occurs in the transition from sleep to waking. (B) Expansion of detail of rhythmic burst firing. (C) Expansion of detail of tonic firing. *From McCormick and Pape (1990).*

current typically has very slow kinetics, turning on with a time constant on the order of tens of milliseconds to seconds. Because the channels underlying this current allow the passage of both Na^+ and K^+ ions, the reversal potential of I_h is typically about −35 mV—between E_{Na} and E_K. Because this current is activated by hyperpolarization below approximately −60 mV, it is typically dominated by the inward movement of Na^+ ions and is therefore depolarizing. For what purpose could neurons use a depolarizing current that activates when the cell is hyperpolarized? A clue comes from cardiac cells in which this current, known as I_f for "funny," is important for determining heart rate. Activation of I_f results in a slow depolarization of the membrane potential between adjacent cardiac action potentials. The more that I_f is activated, the faster the membrane depolarizes between beats, and therefore the sooner the threshold for the next action potential is reached and the next beat is generated. In this manner, the amplitude, or sensitivity to voltage, of I_f can modify the heart rate. Interestingly, the sensitivity of I_f to voltage is adjusted by the release of noradrenaline and acetylcholine; the activation of adrenoceptors by noradrenaline increases I_f and therefore increases the heart rate, whereas the activation of muscarinic receptors decreases I_f, thereby decreasing the heart rate (see DiFrancesco, 2005). This continual adjustment of I_f results from a "push–pull" arrangement between β-adrenergic and muscarinic cholinergic receptors and is mediated by the adjustment of intracellular levels of cyclic AMP. Indeed, the recent cloning of H channels reveals that their structure is similar to that of cyclic nucleotide-gated channels.

Could I_h play a role in neurons similar to that of I_f in the heart? Possibly. Synchronized rhythmic oscillations in the membrane potential of large numbers of neurons, in some respects similar to those of the heart, are characteristic of the mammalian brain. Oscillations of this type are particularly prevalent in thalamic relay neurons during some periods of sleep, as mentioned earlier. Intracellular recordings from these thalamic neurons reveal that they often generate rhythmic "bursts" of action potentials mediated by the activation of a slow spike that is generated through the activation of the low-threshold, or transient, Ca^{2+} current, I_T (Fig. 5.14). Between the occurrence of each low-threshold Ca^{2+} spike is a slowly depolarizing membrane potential generated by activation of the mixed Na^+–K^+ current I_h, as with I_f in the heart. The amplitude, or

voltage sensitivity, of I_h adjusts the rate at which the thalamic cells oscillate, and, as with the heart, this sensitivity is adjusted by the release of modulatory neurotransmitters. In a sense, the thalamic neurons are "beating" in a manner similar to that of the heart.

Summary

An action potential is generated by the rapid influx of Na^+ ions followed by a slightly slower efflux of K^+ ions. Although the generation of an action potential does not disrupt the concentration gradients of these ions across the membrane, the movement of charge is sufficient to generate a large and brief deviation in the membrane potential. Action potentials are typically initiated in the axon initial segment and the propagation of the action potential along the axon allows communication of the output of the cell to its distant synapses. Neurons possess many different types of ionic channels in their membranes, allowing complex patterns of action potentials to be generated and complex synaptic computations to occur within single neurons.

References

Brown, D. A., & Adams, P. R. (1980). Muscarinic suppression of a novel voltage sensitive K^+ current in a vertebrate neurone. *Nature (London), 283,* 673–676.

Brown, D. A., & Passmore, G. M. (2009). Neural KCNQ (Kv7) channels. *British Journal of Pharmacology, 156,* 1185–1195.

Catterall, W. A. (2000a). From ionic currents to molecular mechanisms, The structure and function of voltage-gated sodium currents. *Neuron, 26,* 13–25.

Catterall, W. A. (2000b). Structure and regulation of voltage-gated Ca^{2+} channels. *Annual Review of Cell and Developmental Biology, 16,* 521–555.

Coetzee, W. A., Amarillo, Y., Chui, J., Chow, A., Lau, D., McCormack, T., et al. (1999). Molecular diversity of K^+ channels. *Annals of the New York Academy of Sciences, 868,* 233–285.

Cole, K. S. (1949). Dynamic electrical characteristics of the squid axon membrane. *Archives des Sciences Physiologiques, 3,* 253–258.

Coombs, J. S., Curtis, D. R., & Eccles, J. C. (1957). The interpretation of spike potentials of motoneurons. *Journal of Physiology, 139,* 198–231.

DiFrancesco, D. (2005). Physiology and pharmacology of the cardiac pacemaker ("funny") current. *Pharmacology and Therapeutics, 107,* 59–79.

Hille, B. (1977). Ionic basis of resting potentials and action potentials. In E. R. Kandel (Ed.), *Handbook of physiology* (Sect. 1, Vol. 1, pp. 99–136). Bethesda, MD: American Physiological Society.

Hodgkin, A. L. (1976). Chance and design in electrophysiology. An informal account of certain experiments on nerve carried out between 1934 and 1952. *Journal of Physiology (Landon), 263,* 1–21.

Hodgkin, A. L., & Huxley, A. F. (1939). Action potentials recorded from inside a nerve fiber. *Nature (London), 144,* 710–711.

Hodgkin, A. L., & Huxley, A. F. (1952a). Currents carried by sodium and potassium ions through the membrane of the giant axon of *Loligo*. *Journal of Physiology (London), 116,* 449–472.

Hodgkin, A. L., & Huxley, A. F. (1952b). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology (London), 117,* 500–544.

Huguenard, J., & McCormick, D. A. (1994). *Electrophysiology of the Neuron. An Interative Tutorial*. Oxford: Oxford University Press.

Lauger, P. (1991). *Electrogenic ion pumps*. Sunderland, MA: Sinauer.

Llinás, R. R. (1988). The intrinsic electrophysiological properties of mammalian neurons: Insights into central nervous system function. *Science, 242,* 1654–1664.

McComas, A. J. (2011). *Galvani's spark*. Oxford: Oxford University Press.

McCormick, D. A. (1992). Neurotransmitter actions in the thalamus and cerebral cortex and their role in neuromodulation of thalamocortical activity. *Progress in Neurobiology, 39,* 337–388.

McCormick, D. A., & Pape, H. -C. (1990). Properties of a hyperpolarization-activated cation current and its role in rhythmic oscillation in thalamic relay neurones. *Journal of Physiology (London), 431,* 291–318.

Nicoll, R. A. (1988). The coupling of neurotransmitter receptors to ion channels in the brain. *Science, 241,* 545–551.

Popovic, M. A., Foust, A. J., McCormick, D. A., & Zecevic, D. (2011). The spatio-temporal characteristics of action potential initiation in layer 5 pyramidal neurons: A voltage imaging study. *Journal of Physiology, 589,* 4167–4187.

Reithmeier, R. A. F. (1994). Mammalian exchangers and co-transporters. *Current Opinion in Cell Biology, 6,* 583–594.

Shu, Y., Hasenstaub, A., Duque, A., Yu, Y., & McCormick, D. A. (2006). Modulation of intracortical synaptic potentials by presynaptic somatic membrane potential. *Nature, 441,* 761–765.

Stuart, G., Spruston, N., Sakmann, B., & Hausser, M. (1997). Action potential initiation and backpropagation in neurons of the mammalian CNS. *Trends in Neurosciences, 10,* 125–131.

Suggested Readings

Hille, B. (2001). *Ionic channels of excitable membranes* (3rd ed.). Sunderland, MA: Sinauer.

Hodgkin, A. L. (1992). *Chance and design*. Cambridge: Cambridge Univ. Press.

Huguenard, J., & McCormick, D. A. (1994). *Electrophysiology of the neuron*. New York: Oxford Univ. Press.

Johnston, D., & Wu, S. M. -S. (1995). *Foundations of cellular neurophysiology*. Cambridge, MA: MIT Press.

Koch, C. (1999). *Biophysics of computation*. New York: Oxford Univ. Press.

David A. McCormick

CHAPTER

6

Neurotransmitters

The pioneering studies of Santiago Ramon y Cajal revealed that each cell (neuron) of the brain is an independent structure in close proximity but not physical contact with another neuron (see Shepherd, 1991). Although a half century elapsed before the introduction of the electron microscope led to the final confirmation of this hypothesis, a general acceptance of neurons as the independent building blocks of the brain emerged much more quickly. In turn, this resulted in a new debate: How do these neurons communicate? The answer to this question has evolved continuously over the past century and has become increasingly complex, but the major mode of communication is by chemical messengers.

Neurons share certain structural characteristics (see Chapter 3). They have a cell body (soma) from which processes emerge. One of these types of processes, the axon, is often long but can also be short and thus exert local effects or distant ones. Moreover, axons can branch extensively or have a very confined arbor. In contrast, dendrites are exclusively local structures. The identification of long (axon) and short (dendrite) processes led to the idea that the long axonal projections served to transmit information between neurons over (relatively) long distances, and dendrites were the receptive elements of neurons, responding to the information conveyed by long axons. The gap over which information between neurons must be relayed was termed the *synapse* by Charles Sherrington and came to identify pre- and postsynaptic cells. This general conceptual framework remains in place today, although there are many exceptions, including dendrites that release neuroactive substances and axons that receive inputs from other neurons. One other characteristic proposed by Sherrington that is central to the concept of chemical communication between neurons is that synaptic transmission does not follow all-or-none rules, but is graded in strength and is flexible.

SEVERAL MODES OF NEURONAL COMMUNICATION EXIST

Over the first half of the twentieth century, there was a vigorous debate on the nature of neuronal communication: as one overview of this debate described, it was a war of sparks (electrical communication) and soups (chemical messengers) (see Valenstein, 2005). The debate raged despite the fact that some evidence was marshaled in support of chemical transmission in the mid-nineteenth century. In 1849, Claude Bernard noted that curare, the active constituent of a poison that was applied to arrows in South America, blocked nerve-to-muscle signaling. This effect subsequently was shown to be due to the binding of curare to postsynaptic (muscle) receptors for acetylcholine (ACh), thus blocking neuro-muscular transmission. About 50 years later, Thomas Elliott found that epinephrine caused the contraction of smooth muscle that had been deprived of its nerve inputs, suggesting that muscle contraction depended on the action of chemical molecules liberated from nerves. In a key series of studies using isolated frog hearts, Otto Loewi provided firm evidence in support of chemical neurotransmission by showing that ACh was released upon nerve stimulation and activated a target muscle.

These and other data suggested that the major means of interneuronal communication is chemical in nature, but that neurons also use other processes for intercellular communication. Among these are electrical synaptic transmission, ephaptic interactions, and autocrine, paracrine, and long-range signaling, to which molecules produced by both neural and nonneural cells contribute. The nonsynaptic mode of intercellular communication with the longest range (distance) is hormonal signaling. For example, some hormones made outside of the brain can enter the central nervous system (CNS) to exert effects

© 2013 Elsevier Inc. All rights reserved.

on neurons that express receptors for the hormones. These hormone actions may occur over the short term (changes in neuronal activity) or, more often, over a longer period (long-lasting changes in gene expression).

Molecules released from neurons can also be used in intercellular signaling that does not require synaptic specializations. A variety of molecules are secreted by neurons or diffuse passively from cells, ranging from conventional neurotransmitters to gases such as nitric oxide. These factors may act through autocrine mechanisms (activating receptors on the same cell that releases them) or paracrine pathways to influence nearby cells. Some of these compounds can be retrograde signaling molecules, providing chemically coded feedback to the presynaptic neuron and reversing the more common forward (anterograde) direction of information flow. The role of such molecules is thought to be primarily in modulating neural activity, although they may also provide guidance cues for neurons that are growing toward their final targets in the brain, as well as for the establishment and maintenance of synaptic connections (see Chapters 16 and 18).

CHEMICAL TRANSMISSION

Chemically mediated transmission is the major mode of neuronal communication. The general acceptance of chemical neurotransmission as the means of conveying information between neurons resulted in specific criteria being required for designation of a compound as a neurotransmitter. Based on these "classical" criteria, a relatively small number of compounds have been designated as neurotransmitters. However, over the past generation it has become apparent that a large number of chemical messengers broadly qualify as intercellular transmitters, although these compounds often do not meet—and even seem to sharply differ from—the classical criteria.

Criteria for Designation as a Neurotransmitter

Neurotransmitters are usually considered to be endogenous substances that are released from neurons, act on receptor sites that are typically present on membranes of postsynaptic cells, and produce a functional change in the properties of the target cell. Over the years general agreement evolved that several criteria should be met for a substance to be designated a neurotransmitter:

1. A neurotransmitter must be synthesized by and released from neurons. Thus, the presynaptic neuron should contain both the transmitter and the appropriate enzymes needed to synthesize that neurotransmitter.
2. The substance should be released from nerve terminals in a chemically or pharmacologically identifiable form. It should therefore be possible to isolate the transmitter and characterize its structure.
3. A neurotransmitter should reproduce at the postsynaptic cell the specific events (such as changes in membrane properties) that are seen after stimulation of the presynaptic neuron.
4. The effects of a putative neurotransmitter should be blocked by competitive antagonists of the receptor for that transmitter in a dose-dependent manner. In addition, treatments that inhibit synthesis of the transmitter should block the effects of presynaptic stimulation.
5. There should be active mechanisms to terminate the action of the neurotransmitter. Among such mechanisms are uptake of the transmitter by the presynaptic neurons or glial cells through specific transporter molecules and enzymatic inactivation of the chemical messenger.

The Five Steps of Chemical Neurotransmission

Synaptic transmission consists of a number of steps. The general mechanisms of chemical synaptic transmission are depicted in Figure 6.1.

1. *Synthesis of the neurotransmitter in the presynaptic neuron.* In order for the transmitter to be synthesized, precursors should be present in appropriate places within neurons. Enzymes that convert the precursor(s) into the transmitter should be present in an active form and localized to the appropriate compartment in the neuron. For example, one would not expect synthetic enzymes for the transmitter to be found in the nucleus or the mitochondrion, but these enzymes should be found in the cytosol or in intracellular structures that store the transmitter. In addition, any cofactors that are required for enzyme activity should be present. Drugs that affect the synthesis of neurotransmitters have long been of value in medicine. An example is "α-methyl-*p*-tyrosine, a drug used to treat an adrenal gland tumor that sharply increases blood pressure by releasing very large amounts of norepinephrine (NE); α-methyl-*p*-tyrosine prevents the synthesis of NE and thereby lowers blood pressure.
2. *Storage of the neurotransmitter and/or its precursor in the presynaptic nerve terminal.* Classical and peptide transmitters are stored in synaptic vesicles, where they are sequestered and protected from enzymatic degradation and are ready for quick release. In the case of classical neurotransmitters such as acetylcholine, the synaptic vesicles are small (~50 nm in diameter), in contrast to the large dense-core

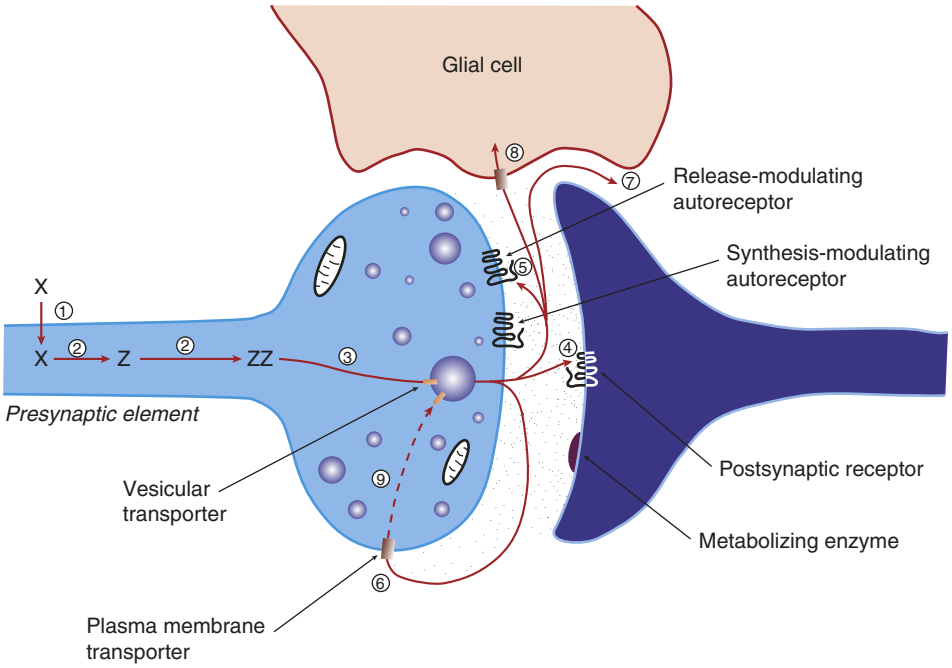


FIGURE 6.1 Schematic representation of the life cycle of a classical neurotransmitter. After accumulation of a precursor amino acid into the neuron (1), the amino acid precursor is metabolized sequentially (2) to yield the mature transmitter. The transmitter is then accumulated into vesicles by the vesicular transporter (3), where it is poised for release and protected from degradation. Once released, the transmitter can interact with postsynaptic receptors (4) or autoreceptors (5) that regulate transmitter release, synthesis, or firing rate. Transmitter actions are terminated by means of a high-affinity membrane transporter (6) that usually is associated with the neuron that released the transmitter. Alternatively, transmitter actions may be terminated by diffusion from the active sites (7) or accumulation into glia through a membrane transporter (8). When the transmitter is taken up by the neuron, it is subject to metabolic inactivation (9).

vesicles (~100 nm in diameter) in which neuropeptide transmitters are stored. Vesicles are often found adjacent to the presynaptic membrane, where they are poised to be released in response to stimulation of the neuron. Because most neurotransmitters are synthesized in the cytosol of neurons, there must be some mechanism through which the transmitter enters the vesicle. This occurs through a vesicular transporter protein that accumulates the transmitter through an energy-dependent process.

3. *Release of the neurotransmitter into the synaptic cleft.* The vesicle in which the transmitter is stored fuses with the cell membrane and releases the transmitter. Neurons use two pathways to secrete proteins. The release of most neurotransmitters occurs by a regulated pathway that is controlled by extracellular signals. The neurotransmitter release process is discussed more fully in Chapter 7. A second (constitutive) pathway of release is not triggered by extracellular stimulation and is used to continuously secrete membrane components, viral proteins, and extracellular matrix molecules. Some unconventional transmitters (such as growth factors) may be synthesized and released by both constitutive and regulated pathways.
4. *Binding and recognition of the neurotransmitter by target receptors.* Once released, neurotransmitters interact with receptors located on the target cell. Most transmitter receptors fall into two broad classes. The first are membrane proteins called metabotropic receptors, which are coupled to intracellular G proteins as effectors (see Chapter 8). Ionotropic receptors form channels through which ions such as Na^+ and Ca^{2+} can enter the neuron. Receptors are also found on the neuron that releases the transmitter. This presynapatic receptor can respond to the transmitter released from the same cell as part of a feedback process to the presynaptic neuron. Such receptors are called autoreceptors, and regulate transmitter release, synthesis, or impulse flow. Alternatively, presynaptic receptors can respond to a chemical signal elaborated from the postsynaptic neuron, another type of feedback mechanism. Finally, receptors on axon terminals can respond to transmitter released from the axon of a different cell, in which case the receptor is called a heteroceptor.
5. *Termination of the action of the released transmitter.* If a cell cannot stop the continued actions of a neurotransmitter, chaotic information flow and sometimes overt cell damage results. Sustained activation of postsynaptic targets can result in tetanus (sustained muscle contraction) or seizures. If one thinks of chemical transmission as being the flow of information, continuous unregulated transmitter release is not conveying to the target cell temporally specific data, such as the rate of firing of the presynaptic neuron or the pattern of firing of the neuron. Neurotransmitter actions may be terminated actively or passively. Among the active termination processes is reuptake of the neurotransmitter through specific transporter proteins on the presynaptic neuron or on glial cells. Another common means of terminating the action of a transmitter is by

enzymatic degradation to an inactive substance. Finally, inactivation can happen by simple diffusion of the transmitter from the synaptic region. All three processes often work cooperatively to terminate the action of a neurotransmitter.

These five steps form a logical scaffold for understanding classical and peptide chemical neurotransmission. There are, however, differences across the various transmitters. We discuss in detail one class of neurotransmitters, the catecholamines, to illustrate the process of chemical neurotransmission, and then examine more briefly other classical neurotransmitters. We also discuss differences between classical and nonclassical transmitters or chemical messengers, including peptide transmitters and unconventional transmitters (such as nitric oxide, growth factors, and endocannabinoids).

CLASSICAL NEUROTRANSMITTERS

The term *classical* is used to differentiate acetylcholine, the biogenic amines, and the amino acid transmitters from other transmitters. The designation is somewhat arbitrary, although these compounds were all accepted as neurotransmitters by the late 1950s. Several factors differentiate classical and other transmitters. First, storage vesicles for classical transmitters are smaller. Second, an energy-dependent reuptake process accumulates classical transmitters that have been released by the presynaptic neuron, allowing subsequent reuse of the transmitter. In contrast, there is no energy-dependent, high-affinity reuptake process for nonclassical transmitters. Third, most classical transmitters are synthesized in the nerve terminal by the actions of one or more enzymes, while peptides are synthesized in the soma from a precursor protein and are then transported to the nerve terminal. The remainder of this chapter deals with the biochemistry of several different transmitter groups, discussing their synthesis, storage, and release, and concludes with new developments that have recently been discovered and expand significantly our understanding of neuronal communication. Other aspects of chemical transmission, including detailed discussions of transmitter release and the receptors for various transmitters is presented in Chapters 7 and 8.

Catecholamine Neurotransmitters

Catecholamines are organic compounds that contain a catechol nucleus (a benzene ring with two adjacent hydroxyl substitutions) and an amine group. The term *catecholamine* is normally used to refer to the three transmitters dopamine (DA), norepinephrine (NE), and epinephrine (Epi). These three neurotransmitters derived from a common precursor are formed by successive enzymatic steps requiring separate enzymes (Fig. 6.2). The presence of these specific synthesizing enzymes in different cells results in different groups of neurons in the brain that synthesize and use as a transmitter either DA, NE, or Epi.

Catecholamines also have transmitter roles in the peripheral (autonomic) nervous system or have certain hormonal functions. For example, catecholamine-containing cells are present in the lining of the intestines where they regulate gut motility. NE is the transmitter of the sympathetic nervous system in mammals, whereas Epi is the sympathetic transmitter in frogs. Despite this species difference, biochemical aspects of neurotransmission are remarkably constant across different vertebrate species and even invertebrates.

Biosynthesis of Catecholamines

The amino acids phenylalanine and tyrosine are precursors for catecholamines. Both amino acids are found in high concentrations in the plasma and brain. In mammals, tyrosine can be formed from dietary phenylalanine by the enzyme phenylalanine hydroxylase, which is found in large amounts in the liver. Insufficient amounts of phenylalanine hydroxylase result in phenylketonuria, a metabolic disorder that leads to intellectual deficits unless treated by dietary manipulation.

Catecholamine synthesis in the brain is usually considered to begin with tyrosine. The enzyme tyrosine hydroxylase (TH) converts the amino acid L-tyrosine into 3,4-dihydroxyphenylalanine (L-DOPA). In turn, L-DOPA is metabolized by L-aromatic amino acid decarboxylase (AADC; see Iversen, Iversen, Bloom, & Roth, 2008) to the transmitter dopamine. This latter step occurs so rapidly that it is almost impossible to measure L-DOPA in the brain without first inhibiting AADC. In neurons that use DA as a transmitter, the decarboxylation of L-DOPA to DA is the final step in transmitter synthesis. However, in neurons that use norepinephrine (also known as noradrenaline) or epinephrine (adrenaline) as transmitters, the enzyme dopamine-β-hydroxylase (DBH), which converts DA to yield NE, is also present. In still other neurons in which epinephrine is the transmitter, a third enzyme (phenylethanolamine *N*-methyltransferase, PNMT) converts NE into Epi. Thus, a cell that uses Epi as its transmitter contains four key enzymes (TH, AADC, DBH, and PNMT), whereas NE neurons contain three enzymes (lacking PNMT), while DA cells have but two transmitter biosynthetic enzymes (TH and AADC).

Tyrosine Hydroxylase

In human beings, a single TH gene is alternatively spliced to yield four TH mRNAs, leading to four TH protein isoforms. In most other primates, only two TH

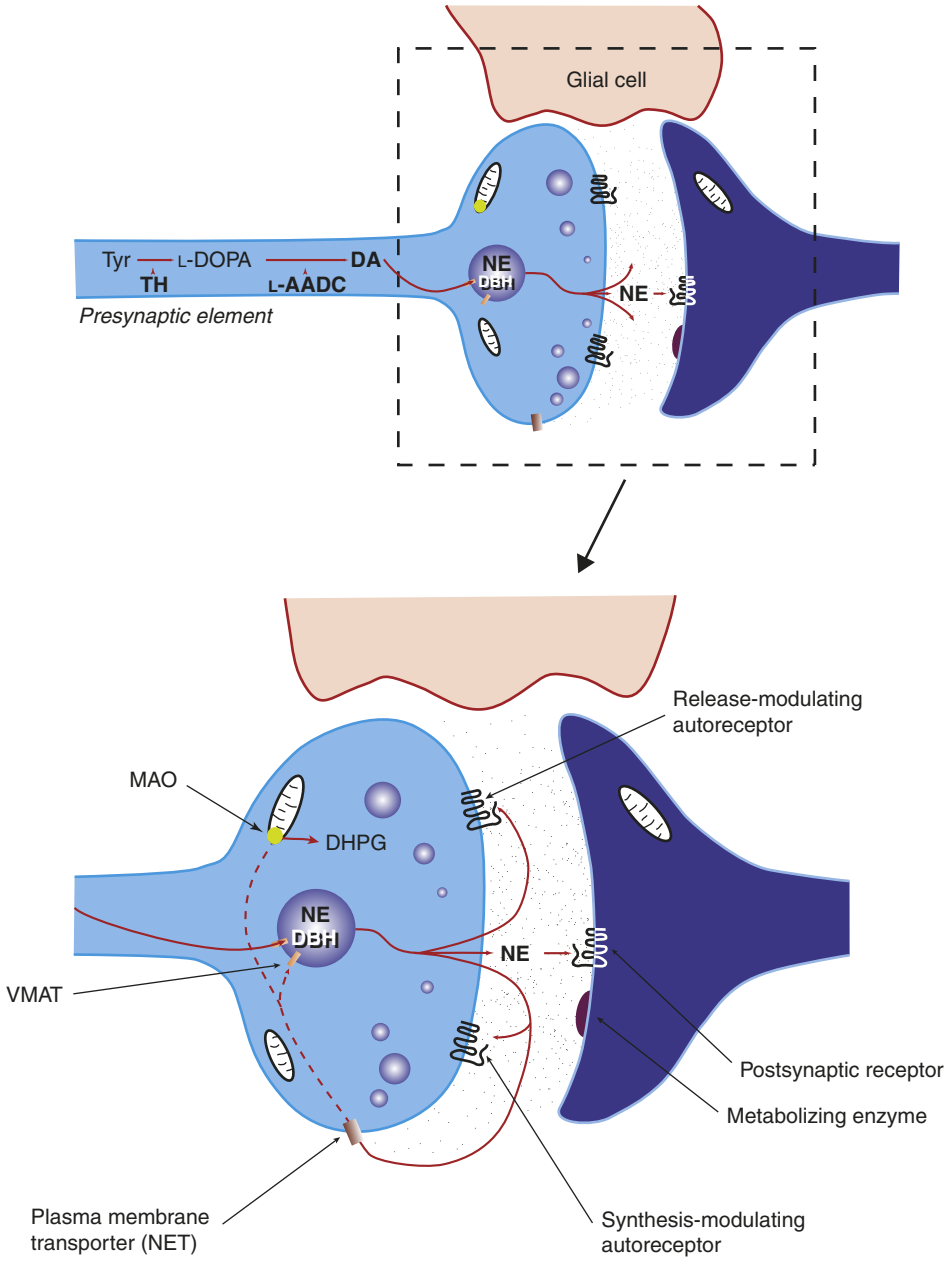


FIGURE 6.2 Characteristics of a norepinephrine (NE)-containing catecholamine neuron. Tyrosine (Tyr) is accumulated by the neuron and then is metabolized sequentially by tyrosine hydroxylase (TH) and L-aromatic amino acid decarboxylase (L-AADC) to dopamine (DA). The DA is then taken up through the vesicular monoamine transporter into vesicles. In DA neurons, this is the final step. However, in this NE-containing cell, DA is metabolized to NE by dopamine-b-hydroxylase (DBH), which is found in the vesicle. Once NE is released, it can interact with postsynaptic noradrenergic receptors or presynaptic noradrenergic autoreceptors. The accumulation of NE by the high-affinity membrane NE transporter (NET) terminates the actions of NE. Once taken back up by the neuron, NE can be metabolized to inactive compounds (DHPG) by degradative enzymes such as monoamine oxidase (MAO) or taken back up by the vesicle.

isoforms are present, and rodents have but a single form of TH. It has been suggested that the human TH isoforms are differentially regulated and have different activities.

TH is a mixed-function oxidase with moderate substrate specificity, hydroxylating phenylalanine, as well as tyrosine. The actions of TH require the cofactor tetrahydrobiopterin (BH4) and iron (Fe^{2+}). BH4 is also an essential cofactor for tryptophan hydroxylase, an enzyme that synthesizes another transmitter, serotonin. Because BH4 is not present in saturating concentrations under basal conditions, it is crucial in regulating TH activity. Intracellular levels of BH4 are determined by its own synthesizing enzyme, GTP cyclohydrolase. Mutations in the gene encoding GTP cyclohydrolase cause DOPA-responsive dystonia, a movement disorder in which DA synthesis is markedly decreased; treatment consists of bypassing the disruption of TH synthesis by administering L-DOPA, the immediate precursor of dopamine.

The two major processes that determine the function of enzymes are mechanisms that control enzyme activity (the rate at which the enzyme converts a precursor into its product) and the amount of newly synthesized enzyme protein. The major factor that determines the activity of TH is phosphorylation of the enzyme at four key serine residues. A second means of regulating enzyme activity is through end-product inhibition: catecholamines can

inhibit the activity of TH by competing for a required cofactor for the enzyme (Iversen et al., 2008).

When neurons require additional catecholamine transmitter to keep demand with release, this can be accomplished by synthesizing new TH protein or increasing the activity of the TH activity. The degree to which increases in catecholamine synthesis depend on *de novo* synthesis of new enzyme protein or changes in the activity of existing enzymes differs across brain regions. For example, chronic increases in the activity of most norepinephrine neurons are accompanied by increases in TH gene expression and subsequent translation of the enzyme. But midbrain dopamine neurons do not appear to meet increased neurotransmitter demand by making more of the enzyme, instead primarily relying on increasing activity of the enzyme through phosphorylation.

The synthesis of catecholamines starts with the entry of tyrosine into the brain. This process is an energy-dependent one in which tyrosine competes with large neutral amino acids as a substrate for a transporter. Because brain tyrosine levels are high enough to saturate TH, catecholamine synthesis cannot usually be increased by administration of tyrosine. Exceptions to this rule are catecholamine synthesis in cells that have a very high activity and thus release the catecholamine transmitter at rapid rates, or in disease states in which entry of tyrosine into the brain is retarded.

L-Aromatic Amino Acid Decarboxylase

As noted above, the hydroxylation of tyrosine by TH generates L-DOPA, which is then decarboxylated to dopamine by the enzyme L-aromatic amino acid decarboxylase (AADC; also sometimes referred to as DOPA decarboxylase). AADC has low substrate specificity and decarboxylates tryptophan as well as tyrosine. Because AADC is found in serotonin as well as catecholamine neurons, it plays an important role in the synthesis of both serotonin and catecholamines. In dopaminergic neurons, AADC is the final enzyme of the synthetic pathway.

Administration of L-DOPA is the primary treatment for Parkinson's disease, which is due to loss of dopamine. Dopamine itself cannot readily enter the brain, but the dopamine precursor L-DOPA freely enters the brain and, after being accumulated by dopamine neurons, is metabolized to dopamine. Although L-DOPA crosses the blood–brain barrier to enter the brain, systemically administered L-DOPA is so rapidly metabolized in the blood that little of the precursor actually reaches the brain. To circumvent this problem L-DOPA is administered together with a second drug that does not enter the brain but inhibits decarboxylases in the bloodstream.

Dopamine-β-Hydroxylase

Noradrenergic and adrenergic neurons contain the enzyme dopamine-β-hydroxylase (DBH), which converts DA into NE. In noradrenergic neurons, this is the final step in catecholamine synthesis. Humans have two different DBH mRNAs that are generated from a single gene. DBH has poor substrate specificity and *in vitro* oxidizes almost any phenylethylamine to its corresponding phenylethanolamine. For example, in addition to forming NE from DA, DBH converts tyramine into octopamine. Receptors that have a high affinity for the trace amines tyramine and octopamine have recently been discovered (see Zucchi, Chiellini, Scanlan, & Grandy, 2006) and may be involved in certain psychiatric disorders and in the side effects of certain drugs used to treat major depression.

Phenylethanolamine-N-Methyltransferase

Phenylethanolamine-N-methyltransferase is present at high levels in a small number of brainstem neurons that use epinephrine as a transmitter. PNMT is also found in the inner portion (medulla) of the adrenal gland, where it methylates NE to form Epi, the major adrenal catecholamine. There is a single PNMT gene. Brain PNMT regulation has not been extensively studied, but expression of the mRNA encoding PNMT is regulated by adrenal glucocorticoids and nerve growth factor.

Storage of Catecholamines and Their Enzymes

Catecholamines are stored in specialized subcellular organelles called vesicles. Accordingly, a specialized vesicular transporter protein is required in order for dopamine and other transmitters that are formed in the cytosol to enter the vesicle (El Mestikawy, Wallén-Mackenzie, Fortin, Descarries, & Trudeau, 2011).

Vesicular Storage

The vesicle serves as a depot for the transmitter until it is released by appropriate physiological stimuli (see Chapters 5 and 7). In addition, by storing transmitters in an organelle, vesicles protect the transmitter from metabolic inactivation by cytosolic enzymes.

The norepinephrine-synthesizing enzyme DBH differs from the other enzymes in catecholamine synthesis by being found in the vesicle rather than the cytosol. This means that only *after* DA is taken up by vesicles through a vesicular monoamine transporter (VMAT) is it metabolized to NE. This vesicular storage of DBH has an interesting consequence: DBH and dopamine can both be released into the synapse along with NE.

Vesicular Monoamine Transporters

The ability of vesicles to accumulate DA and other compounds depends on VMAT. The vesicular transporter differs from the neuronal membrane transporter that terminates the action of catecholamines in terms of substrate affinity and localization. Two VMAT genes have

been cloned: VMAT1 is found in adrenal gland cells that synthesize and release catecholamines, and VMAT2 is found in the catecholamine and serotonin-containing neurons of the brain. VMAT2 is not very specific and accumulates catecholamines, indoleamines such as serotonin, and histamine into vesicles.

The VMATs are inhibited by reserpine. This drug has been used for centuries in India as a folk medicine to treat high blood pressure and psychoses. These uses of reserpine were reported in international journals in the early 1930s, but these therapeutic actions were not appreciated in Western medicine until a generation later; the unpleasant side effects of reserpine initially garnered more attention. In the 1950s Bernard Brodie and coworkers discovered that reserpine depleted brain serotonin. The contemporaneous discovery that the hallucinogen LSD is structurally similar to serotonin led to the proposal that the antipsychotic actions of reserpine were due to its ability to deplete serotonin in the brain. However, it was soon realized that reserpine depletes both serotonin and catecholamines in the brain; thus, antipsychotic effects of reserpine might be due to serotonin or catecholamine depletion, or both.

To determine if serotonin or catecholamines were more important for the antipsychotic effect of reserpine, Arvid Carlsson and colleagues administered catecholamine and serotonin precursors to rats treated with reserpine in an attempt to replenish the levels of the transmitters. At the time of these studies dopamine was not thought to be a transmitter, but was considered to be present solely as a precursor of norepinephrine. Carlsson and collaborators examined locomotor activity, which is severely depressed by reserpine. Motor function was restored by the DA precursor L-DOPA but not by the serotonin precursor 5-hydroxytryptophan. This work led to the characterization of DA as a neurotransmitter. The demonstration that L-DOPA increases brain DA concentrations in animals treated with reserpine to deplete vesicular stores of DA also suggested that the primary mechanism through which reserpine exerts antipsychotic effects was through its ability to disrupt DA transmission. This idea led directly to the hypothesis that the dysfunction of central DA systems underlies schizophrenia.

VMATs have significant homology with bacterial antibiotic drug resistance transporters, suggesting a role of VMAT in detoxification. This is indeed the case: VMAT can sequester toxins (as it sequesters transmitters) in vesicles. This is perhaps best illustrated by studies of genetically modified mice with low VMAT2 expression. In these mice, the toxin MPTP (which disrupts mitochondrial function and elicits dopamine neuron degeneration) results in a greater loss of dopamine than seen in control mice. This is because less MPTP is sequestered in vesicles, with more MPTP in the cytosol where it may disrupt mitochondria.

Release of Catecholamines

Catecholamine release usually occurs through a calcium-dependent exocytotic process involving brief fusion of the transmitter-containing vesicle with the plasma membrane (see Chapter 8). However, catecholamine release can also occur through at least two other mechanisms. First, catecholamines can be released by a reversal of the dopamine and NE transporters to extrude the catecholamines; this mode of release is seen after administration of amphetamine. Second, dopamine can be released from dendrites through a process that does not appear to involve calcium.

Regulation of Catecholamine Synthesis and Release by Autoreceptors

Enzymes that control catecholamine synthesis can be regulated at both the transcriptional level and by post-translational modifications (such as phosphorylation) that alter enzymatic activity. In addition, DA and NE synthetic enzymes can be regulated by autoreceptors on the nerve terminal.

Autoreceptors are found on most parts of the neuron and are defined functionally by the events that they regulate. Thus, synthesis-, release-, and impulse-modulating DA autoreceptors have been described (see Iversen et al., 2008). All three types of DA autoreceptors belong to the D2 family of DA receptors, which includes three different receptors (D_2, D_3, and D_4). It is clear that D_2 autoreceptors exist, but the presence of D_3 DA autoreceptors remains controversial; there are no D_4 autoreceptors. Because all three types of autoreceptors are D_2 receptors, there must be different coupling to distinct intracellular signaling cascades to acount for the functional differences between release-, synthesis, and impulse-modulating autoreceptors.

The autoreceptor can be thought of as a key part of a feedback mechanism. Thus, DA that is released from a neuron stimulates an autoreceptor to inhibit further DA release. *Release-modulating* autoreceptors are a common regulatory feature of all neurons that use classical transmitters. Because *intra*cellular DA levels regulate tyrosine hydroxylase activity by interfering with the binding of TH to its cofactor, changes in the release of DA may also alter transmitter synthesis. *Synthesis-modulating* autoreceptors directly regulate DA synthesis: DA release and subsequent binding to the D_2 autoreceptor decreases DA synthesis, while DA antagonists that block the autoreceptor increase catecholamine synthesis. Interestingly, synthesis-modulating autoreceptors are not found on all DA neurons: some midbrain and hypothalamic DA neurons lack synthesis-modulating autoreceptors. Finally, *impulse-modulating* autoreceptors are found on the soma and dendrites of DA neurons and regulate the firing rates of these cells. Because the release of DA can alter synthesis

of the transmitter by changing feedback inhibition on TH, it follows that impulse-modulating autoreceptors also change DA synthesis. Thus, all three types of DA autoreceptors may ultimately regulate synthesis. The interdependence of regulatory processes governing DA cells is characteristic of monoaminergic neurons.

Inactivation of Catecholamine Transmission

Constant levels of a neurotransmitter in the synapse will not accurately convey information about the dynamic state of the presynaptic neuron to its follower cell. In addition, continuous stimulation of certain receptors is pathological and can damage postsynaptic cells. For example, when certain receptors for the transmitter glutamate are activated continuously, the result is "excitotoxic" cell death. In order to prevent continuous presynaptic signaling, different mechanisms have evolved for terminating the actions of a transmitter. The simplest is diffusion of the transmitter and its subsequent dilution in extracellular fluid to subthreshold concentrations. More important, however, are active modes of halting transmitter action, including uptake of the transmitter by membrane-associated transporter proteins and enzymatic inactivation of the transmitter.

Enzymatic Inactivation of Catecholamines

Enzymatic inactivation originally was thought to be the major means of terminating catecholamine actions in the CNS. Two enzymes contribute to catecholamine catabolism: monoamine oxidase (MAO) and catechol-*O*-methyltransferase (COMT). These enzymes can act independently or can act on the products generated by the other enzyme, leading to catecholamine metabolites that are deaminated, *O*-methylated, or both. COMT is a relatively nonspecific enzyme that transfers methyl groups from the donor *S*-adenosylmethionine to the *m*-hydroxy group of catechols. COMT is found both in peripheral tissues and the central nervous system and is the major means of inactivating catecholamines that are released from the adrenal gland.

MAOs oxidatively deaminate catecholamines and their *O*-methylated derivatives to form inactive and unstable derivatives that can be further degraded by other enzymes. Two forms of MAO have been identified. MAO_A has high affinities for NE and serotonin and is selectively inhibited by drugs such as clorgyline. In contrast, MAO_B has a higher affinity for *o*-phenylethylamines and is selectively inhibited by different compounds, such as deprenyl. The MAOs are important targets of drugs used to treat several neuropsychiatric disorders (Box 6.1).

Neuronal Catecholamine Transporters

The reuptake of a transmitter released by a neuron is the major mode of inactivation of classical neurotransmitters in the brain. Accumulation of the transmitter also allows intracellular enzymes that degrade the transmitter to act, thus bolstering the actions of extracellular enzymes.

Several characteristics define the high-affinity reuptake process for transmitters. The process is energy dependent and saturable, depends on Na^+ cotransport, and requires extracellular Cl^-. Because reuptake depends on coupling to the Na^+ gradient across the neuronal membrane, toxins that inhibit Na^+,K^+-ATPase inhibit reuptake. Under certain conditions, the coupling of transporter function to Na^+ flow may cause local changes in membrane Na^+ gradients and thereby paradoxically cause the transporter to operate in "reverse" to "release" the transmitter from the cell to the extracellular space.

Membrane catecholamine transporters are not Mg^{2+} dependent and are not inhibited by reserpine, which distinguishes neuronal and vesicular membrane transporters. Catecholamine transporters are localized to neurons. Although a reuptake process accumulates catecholamines into astrocytes, a type of glial cell, this is not a high affinity process and its functional significance is unclear.

Two distinct mammalian catecholamine transporters, one for dopamine (dopamine transporter, DAT) and one for norepinephrine (NET), have been identified. The two are closely related members of a class of transporter proteins (including serotonin and amino acid transmitter transporters) with 12 transmembrane domains. Neither transporter is specific, with each accumulating both DA and NE. In fact, NET has a higher affinity for dopamine than for NE.

The regional distribution of DAT and NET largely follows the expected localization to DA and NE neurons, respectively. However, DAT does not appear to be expressed in all DA cells. Certain hypothalamic cells that release DA into the blood system of the pituitary lack detectable DAT mRNA and protein. Because DA released from these neurons is carried away in the blood, there is no need for a transporter protein on these DA cells.

Immunohistochemical studies of the subcellular localization of DAT found that the transporter is not present at the synaptic junction, but is located just outside of this region. Thus, the transporter may be used to accumulate DA that has escaped from the synaptic cleft. This suggests that diffusion is the initial process by which DA is removed from the synapse. The extrasynaptic localization of the catecholamine transporters, coupled with a similar extrasynaptic localization of dopamine receptors, suggests that extrasynaptic ("volume") neurotransmission may be of major importance in transmitter actions (see Agnati, Guidolin, Guescini, Genedani, & Fuxe, 2010).

Mice in which the gene encoding DAT has been deleted have been particularly useful in clarifying the function of transporters. Transgenic mice that lack DAT have a remarkable number of changes in DA function, ranging from increased extracellular DA levels and

BOX 6.1

MAO AND COMT INHIBITORS IN THE TREATMENT OF NEUROPSYCHIATRIC DISORDERS

One hypothesis of the pathophysiology of depression posits a decrease in noradrenergic levels in the brain. MAO_A inhibitors, such as tranylcypromine, effectively increase NE levels (as well as DA and 5-HT concentrations) and were once a mainstay in the treatment of depression. However, the use of MAO inhibitors in depression is less common today, with the major pharmacological treatments being drugs that increase extracellular NE levels by blocking the NE transporter (tricyclic antidepressants) and other drugs that increase 5-HT or DA levels by blocking SERT or DAT (such as Prozac and Welbutrin). The treatment of depression with MAO_A inhibitors, although still useful for certain patients who do not respond to other antidepressants, is marred by a large number of side effects. Among the most serious is hypertensive crisis. Patients treated with MAO_A inhibitors cannot metabolize tyramine efficiently, which is present in large amounts in foods such as aged cheeses and red wines. Because tyramine releases catecholamines peripherally, small amounts of tyramine increase blood pressure significantly and increase the risk for a stroke.

Deprenyl, a specific inhibitor of MAO_B, has been used as an initial treatment for Parkinson's disease (PD; see Chapter 31). The use of deprenyl in the treatment of PD and the rationale for its use were based on data from studies of a neurotoxin, 1-methyl-4-pheny 1-1,2,3,6-tetrahydropyridine (MPTP). MPTP causes degeneration of midbrain DA neurons and a parkinsonian syndrome. MPTP-induced parkinsonism was first noted in a group of opiate addicts: in an attempt to synthesize a designer drug, the structurally related MPTP was inadvertently produced. Addicts who injected this drug developed a severe parkinsonian syndrome. Subsequent animal studies showed that MPTP itself is not toxic, but that its active metabolite, MPP^+, is highly toxic. The formation of MPP^+ from MPTP is catalyzed by MAO_B, and thus treatment with MAO inhibitors such as deprenyl can prevent MPTP toxicity. The realization that MPTP administration reproduces the motor deficits on PD reawakened interest in environmental toxins as a cause of PD. The MPTP saga also led to the idea that deprenyl treatment might slow the progression of PD by preventing metabolism of some environmental compound to an active toxin. Although clinical studies initially were interpreted to suggest that deprenyl treatment slowed progression of Parkinson's disease, later studies showed that deprenyl increases DA levels slightly and thus gives some symptomatic relief.

Catechol O-methyl transferase, which together with MAO degrades catecholamines, also plays a role in the treatment of PD. Two COMT inhibitors are used to prevent the enzymatic inactivation of DOPA. By inhibiting COMT, these drugs prolong the therapeutic action of DOPA and may smooth out fluctuations in the therapeutic response to DOPA.

Changes in catecholamine function also have been the object of intense scrutiny in schizophrenia, with recent attention focusing on possible changes in dopamine. One allelic variant of the COMT gene results in a much reduced activity of the enzyme. Data have examined COMT alleles for full vs. low COMT activity in normal subjects and schizophrenics. Individuals bearing the allele that confers lower COMT activity display improved performance on cognitive tasks that involve DA actions in the prefrontal cortex; the performance of schizophrenic persons on these tasks is impaired. It therefore has been proposed that high COMT activity may confer an increased risk to schizophrenia.

Ariel Y. Deutch and Robert H. Roth

delayed clearance of released DA to a striking decrease in tissue concentrations of DA in the face of increased DA synthesis (Gainetdinov, Jones, Fumagalli, Wightman, & Caron, 1998). Some of the deficits in DAT knockout mice have been suggested to reflect a disinhibition of tyrosine hydroxylase due to a lack of intraneuronal DA (removing feedback inhibition of the enzyme), resulting in a marked increased in DA synthesis and release.

The catecholamine transporters are important targets of many drugs. Cocaine and amphetamine both increase extracellular levels of catecholamines by blocking transporters. In particular, cocaine shows a very high affinity for DAT; amphetamine is a less potent inhibitor but also "releases" catecholamine by reversing the normal direction of transporter function, extruding the transmitter from neurons. The tricyclic antidepressant drugs work by blocking the norepinephrine transporter, as do newer norepinephrine-selective uptake blockers.

Serotonin

Well over a century ago, scientists were aware of a substance in the blood that induced powerful contractions of smooth muscle organs. In the mid-twentieth

century, Page and collaborators succeeded in isolating the compound, which they suggested to be a possible cause of high blood pressure, from platelets. At the same time, Italian researchers were studying a substance in intestinal mucosa that caused contractions of intestinal smooth muscle. The material isolated from platelets was called "serotonin," and the substance isolated from the intestinal tract was named "enteramine." Studies soon revealed that the two substances were the same compound, 5-hydroxytryptamine (5-HT), which now is commonly referred to as serotonin.

Serotonin is found in neurons and in several other types of cells in the body. In fact, the brain accounts for only about 1% of total body stores of serotonin. Although the isolation and identification of serotonin were made from peripheral tissues, much of the subsequent interest in serotonin was based on its potential involvement in psychiatric disorders. The finding that serotonin's chemical structure is similar to that of LSD led to theories that associated abnormalities in serotonin function to schizophrenia and depression.

The basic process of serotonin synthesis is very similar to that of catecholamine transmitters: a peripheral amino acid (tryptophan) gains entry into the brain and is metabolized in serotonergic neurons via a series of enzymatic steps. Once tryptophan enters the serotonergic neuron, it is hydroxylated by tryptophan hydroxylase, the rate-limiting step in serotonin synthesis, giving rise to 5-hydroxytryptophan (5-HTP). In turn, 5-HTP is decarboxylated by L-aromatic amino acid decarboxylase to the transmitter serotonin. Thus, only two critical enzymes (tryptophan hydroxylase and AADC) are involved in the synthesis of 5-HT (Iversen et al., 2008).

Alternative Tryptophan Metabolic Pathways

Although serotonin is usually the endpoint of tryptophan metabolism in brain, serotonin can be metabolized further to yield active products. In the pineal gland, 5-HT is metabolized to form 5-methoxy-N-acetyltryptamine (melatonin), a hormone thought to play an important role in sleep.

In peripheral tissues, most tryptophan is not used to synthesize 5-HT but instead is metabolized by the kynurenine pathway. This kynurenine "shunt" is also present in the brain and leads to the accumulation of several interesting active substances (see Schwarcz, 2004). The two major tryptophan metabolites generated by the kynurenine shunt are quinolinic and kynurenic acids. Quinolinic acid is a potent activator of certain glutamate receptors and upon administration to animals causes cell loss and convulsions. In contrast, kynurenic acid is an agonist that activates one type of acetylcholine receptor in brain and decreases extracellular levels of the transmitter glutamate; levels of kynurenic acid have been reported to be increased in schizophrenia.

Inactivation of Released Serotonin

Serotonin is inactivated primarily by reuptake through SERT, the serotonin transporter that belongs to the same family of transporters as the catecholamine transporters. Selective serotonin reuptake blockers (SSRIs), which block SERT, are widely used as antidepressants; the prototypic SSRI is fluoxetine (Prozac). Because different antidepressant drugs increase levels of serotonin or NE, or both, current theories of depression posit critical modulatory roles for NE and 5-HT.

The enzymatic degradation of 5-HT is catalyzed by monoamine oxidase (MAO). The product of this reaction, 5-hydroxyindole acid aldehyde, is oxidized further to 5-hydroxyindoleacetic acid. MAO inhibitors prevent the metabolism of serotonin and thus increase 5-HT levels, and have been widely used as antidepressants (Box 6.1).

Acetylcholine

Our basic ideas about chemical synaptic transmission are based on early studies of acetylcholine (ACh). Pharmacological and electrophysiological studies at the neuromuscular junction led to an understanding that ACh is a chemical messenger between excitable cells. These ideas about cholinergic signaling between cells are the cornerstone of today's understanding of chemical neurotransmission. For example, electrophysiological studies revealed fast excitatory responses of muscle fibers to the stimulation of nerves innervating the muscle. The presence of miniature end plate potentials (mEPPs) in muscle fiber was noted, and in the early 1950s, Fatt and Katz demonstrated that these mEPPs resulted from the slow "leakage" of ACh, with each mEPP representing the release of transmitter in one vesicle (termed a quantum). Overt depolarization generated an increase in the number of quanta released over a given period of time (see Chapter 7). Over the past half century many of the rules that govern ACh neurotransmission have been found to be general principles that apply to many transmitters.

The synthesis of ACh has a single step: the acetyl group from acetyl coenzyme A is transferred to choline by the enzyme choline acetyltransferase (ChAT), forming Ach. The acetyl-CoA that serves as the donor is derived from pyruvate generated by glucose metabolism. There is one interesting twist on ACh synthesis: acetyl-CoA is found in mitochondria, but the synthetic enzyme ChAT is cytoplasmic. This means that acetyl-CoA must exit the mitochondria to gain access to ChAT. Despite the fact that ChAT is the only enzyme involved in ACh synthesis, it is not the rate-limiting step in synthesis of the transmitter. The full enzymatic activity of ChAT is not expressed *in vivo*: ChAT activity measured

in vitro is much greater than would be expected on the basis of measurements of ACh synthesis *in vivo*. The reason for this discrepancy has been suggested to be related to the requirement that acetyl-CoA be transported from the mitochondria to the cytoplasm. Alternatively, intracellular choline concentrations may determine the rate of ACh synthesis. This latter idea led to the use of choline precursors in attempts to increase ACh synthesis in Alzheimer's disease, in which there is a marked decrease of cortical ACh levels. Unfortunately, these attempts have not proven successful.

ACh is taken up into storage vesicles by the vesicular cholinergic transporter (VAChT). Cloning of the human VAChT revealed that the gene is localized to chromosome 10, near the gene encoding ChAT. Subsequent studies demonstrated that the entire VAChT coding region is contained in the first intron of the ChAT gene. These observations suggested that both genes are coordinately regulated, a suspicion that subsequently was confirmed.

A major way of inactivating ACh is enzymatic, in which ACh is hydrolyzed to choline. There are two types of cholinesterases that accomplish this step: acetylcholinesterases (AChEs) and butyrylcholinesterases (Taylor & Brown, 2006). The former are relatively specific for ACh and are found in high concentration in the brain, whereas butyrylcholinesterases are enriched in the liver.

Acetylcholinesterase is present in high concentrations in cholinergic neurons. However, AChE is also found in moderately high concentrations in some noncholinergic neurons that receive cholinergic inputs. This observation is consistent with the fact that AChE is a secreted enzyme that is associated with the cell membrane. Thus, ACh hydrolysis takes place extracellularly, and the choline generated is conserved by the high-affinity reuptake process. In addition to its role in inactivating acetylcholine, AChE has been proposed to be a chemical messenger in the CNS (Greenfield, Zimmermann, & Bond, 2008).

γ-Aminobutyric Acid: The Major Inhibitory Neurotransmitter

Several amino acids fulfill most of the criteria for consideration as neurotransmitters. The best studied of these are γ-aminobutyric acid (GABA), the major inhibitory transmitter in brain, and glutamate, which is the brain's major excitatory transmitter. Many of the principles concerning transmitter synthesis and inactivation that we have previously discussed also apply to amino acid transmitters. However, there are also some key differences. The most obvious difference is that GABA is derived from glucose metabolism (Fig. 6.3). Thus, mechanisms must exist to segregate the transmitter and general metabolic pools of the amino acid transmitters (Olsen & Betz, 2006). A second difference between amino acid and catecholamine transmitters is that inactivation of the former involves several types of transporters that are found on glial cells as well as neurons.

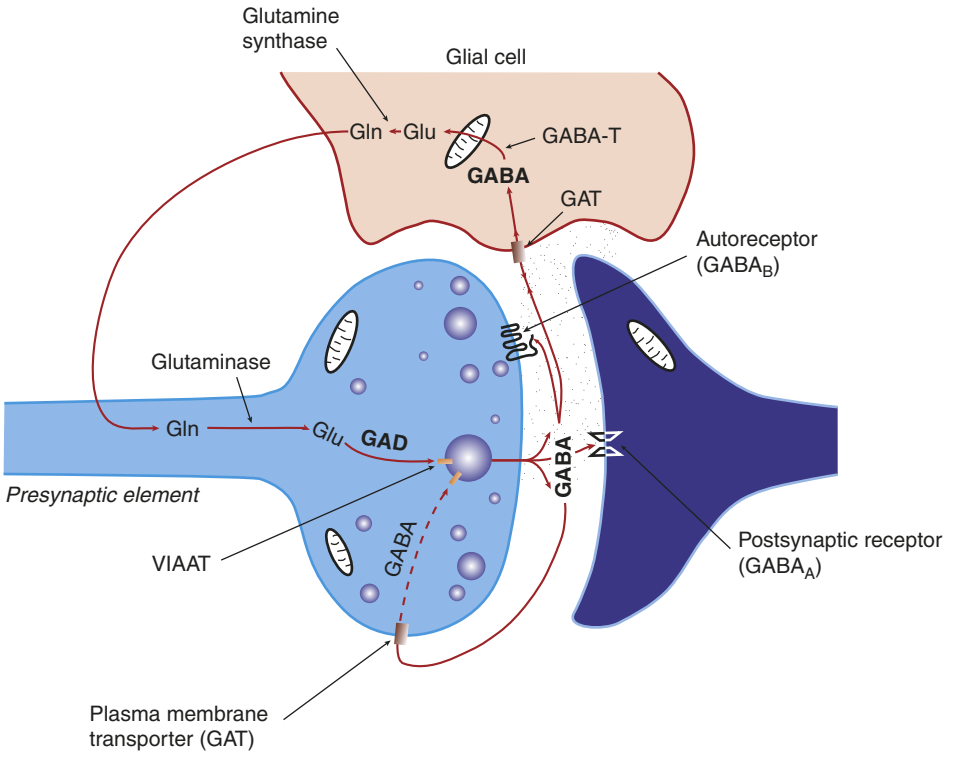


FIGURE 6.3 Schematic depiction of the life cycle of the transmitter pool of GABA. a-Ketoglutarate formed in the Krebs cycle is transaminated to glutamate (Glu) by GABA transaminase (GABA-T). The transmitter GABA is formed from the Glu by glutamic acid decarboxylase (GAD). GABA that is released is taken by high-affinity GABA transporters (GAT) present on neurons and glia.

GABA Synthesis

The GABA Shunt and GABA Transaminase

GABA ultimately is derived from glucose metabolism. α-Ketoglutarate formed by the Krebs cycle is transaminated to the amino acid glutamate by GABA-transaminase (GABA-T). In those cells in which GABA is used as a transmitter, the presence of another enzyme, glutamic acid decarboxylase (GAD), generates GABA from glutamate (Fig. 6.3). The presence of GAD is therefore an unambiguous marker of neurons that use GABA as a transmitter.

An unusual feature of GABA synthesis is that intraneuronal GABA is inactivated by the actions of GABA-T, which appears to be associated with mitochondria. Thus, GABA-T is both a synthetic and a degradative enzyme! GABA-T metabolizes GABA to succinic semialdehyde but only if α-ketoglutarate is present to receive the amino group that is removed from GABA. This unusual GABA shunt serves to maintain supplies of GABA.

How Are Transmitter and Metabolic Pools Separate?

A major question concerning amino acid transmitters is how the transmitter pool is separately maintained from the general metabolic roles of amino acids. The GABA synthetic enzyme GAD is a cytosolic enzyme, but GABA-T, which converts α-ketoglutarate into the GAD substrate glutamate, is present in mitochondria. Thus, the metabolic pool is present in mitochondria, and glutamate destined for the transmitter pool must be exported from mitochondria to the cytosol.

Glutamate is not only a precursor to GABA, but it is also the major excitatory neurotransmitter found in the brains of mammals. Why GABA neurons fail to use the precursor glutamate as a transmitter probably involves different vesicular transporters that accumulate glutamate and GABA into vesicles, as well as different biosynthetic enzymes for the transmitter and metabolic pools of glutamate. In the case of glutamate, three different vesicular glutamate transporters are responsible for determining if glutamate is used as a transmitter.

Storage and Release of GABA

Vesicular Inhibitory Amino Acid Transporter

A vesicular GABA transporter was cloned on the basis of homology to a protein found in the worm *Caenorhabditis elegans*. The strategy of moving from invertebrate to mammalian species has been very useful in identifying a variety of mammalian transmitter-related genes. The vesicular GABA transporter shares with the vesicular monoamine transporters a lack of substrate specificity, and will transport the inhibitory transmitter glycine as well as GABA.

Autoreceptor Regulation of GABA Release

There are two major classes of GABA receptors. One type of GABA receptor (the $GABA_A$ receptor) forms ion channels and is mainly on cells postsynaptic to GABA terminals (see Chapter 8). A second type of receptor, the $GABA_B$ receptor, is coupled to intracellular G proteins. Pharmacological studies indicate that the autoreceptor-mediated regulation of GABA neurons involves $GABA_B$ receptors located on GABAergic axon terminals. Anatomical studies have revealed that $GABA_B$ receptors also are found on postsynaptic non-GABAergic neurons. It is possible that these postsynaptic $GABA_B$ sites respond to GABA released from a neuron that is presynaptic to another GABA neuron. Interestingly, when one GABA neuron terminates on another GABA cell, the inhibition of the second (postsynaptic cell) will have the same functional consequence as an autoreceptor (decreasing subsequent transmitter release) on the third cell in the chain.

Inactivation of GABA

Reuptake is the primary mode of inactivation of the transmitter GABA. There are four GABA transporters (GATs), providing a diverse means of regulating GABA neurons. Early studies defined the different GABA transporters as neuronal or glial based on pharmacological criteria. However, anatomical studies found that one GAT that was defined on pharmacological grounds as a glial transporter is present in both neurons and glia. The reason for multiple GABA transporters is not clear. GATs are expressed in both GABAergic and non-GABAergic cells (presumably cells that receive a GABA innervation). Consistent with the promiscuous uptake of transmitters by other transporters, amino acids other than GABA are also substrates for GATs.

Glutamate and Aspartate: Excitatory Amino Acid Transmitters

Excitatory amino acid transmitters account for most of the fast synaptic transmission that occurs in the mammalian brain. Glutamate and aspartate are the major excitatory amino acid neurotransmitters, and several related amino acids, such as N-acetylaspartylglutamate, are also thought to have neurotransmitter roles.

Neither glutamate nor aspartate crosses the blood–brain barrier, and in the brain these transmitters are derived by local synthesis from glucose. Two processes lead to glutamate synthesis in the nerve terminal. As discussed earlier, glutamate is formed from glucose through the Krebs cycle and transamination of α-ketoglutarate. Glutamate can also be formed directly from glutamine. Because glutamine is synthesized in glial cells, both

neurons and glia are important in determining the transmitter pool of glutamate. Glutamine is exported from glia and incorporated by nerve terminals before being converted into glutamate by a glutaminase enzyme (Hassel & Dingledine, 2006).

Because of the intermingling of glial and neuronal contributions to glutamate synthesis, and the lack of specific enzymes or other proteins to distinguish the metabolic pool of glutamate from the transmitter glutamate, the identification of glutamatergic neurons was initially made on the basis of physiological studies. More recently the identification of three vesicular glutamate transporters, VGluT1-3, has provided a relatively simple means for identifying glutamatergic neurons. However, in neuroscience facts are rarely unambiguous, and anatomical studies over the past several decades have revealed that many glutamate neurons also contain one or more other transmitters (Fremeau, Voglmaier, Seal, & Edwards, 2004), with some even containing GABA. Because GABA hyperpolarizes cells in adult neurons, while glutamate depolarizes neurons, one would not expect to find transmitter pools of GABA and glutamate in the same neuron. The functional significance of such an arrangement is not clear.

Just as there are several vesicular glutamate receptors, there are several glutamate transporters that terminate the action of glutamate. An unusual aspect is that the major cell type taking up glutamate is the astrocyte, which expresses very high levels of one glutamate transporter. This is one of several findings that point to glia as releasing glutamate in a signaling capacity (Halassa & Haydon, 2010). These findings have led to an increasing awareness of glia as being far more than structural support cells and argue for a much broader contribution of glia to neuronal communication.

NONCLASSICAL NEUROTRANSMITTERS

Classical transmitters account for relatively few of all neurotransmitters. For example, many more transmitters are peptides, which share many but not all key attributes of classical transmitters. Both classical and peptide transmitters usually are well conserved across species, and indeed many of the peptide transmitters were initially isolated from amphibians. In addition, both classical and peptide transmitters are synthesized in neurons, where they are stored in vesicles and released in a Ca^{2+}-dependent manner. However, the synthetic mechanisms and the modes of inactivation of the two types of transmitters are quite different. We will first consider the question of the significance of multiple neurotransmitters, and then turn to general principles of peptide transmitter life cycle.

Why Do Neurons Have So Many Transmitters?

About a dozen molecules are generally accepted as "classical" transmitters. Many more neuropeptides serve as transmitters. And there are "unconventional" transmitters that fail to meet most of the criteria for classical transmitters yet have clearly been shown to function as chemical messengers over restricted spatial and temporal domains. If transmitters simply serve as a chemical bridge that conveys information between two spatially distinct cells, why have so many chemical messengers?

Convergence of Different Transmitter-containing Axons on a Common Neuron

Perhaps the simplest explanation for multiple transmitters is that many nerve terminals synapse onto a single neuron. How can a neuron distinguish between multiple inputs that carry different information? One way is to segregate the place on the neuron at which an input arrives, such as the soma, axon, or dendrite. However, because many afferents terminate in close proximity, another means of distinguishing inputs and their information is necessary. Chemical coding of inputs from various neurons can provide a high degree of both spatial and temporal resolution.

Colocalization of Neurotransmitters

The idea that a neuron is limited to one transmitter can be traced to Henry Dale, or more accurately, to an informal restatement of what is termed Dale's principle. In the 1940s, Dale posited that a metabolic process that takes place in the cell body can reach or influence events in all processes of the neuron. Sir John Eccles restated Dale's view to suggest that a neuron releases the same transmitter from all of its processes. Illustrating the dangers of scientific sound bites, this "principle" was soon misinterpreted to indicate that only a single transmitter can be present in a given neuron, a view that held sway for about 40 years. We now know that neurons can contain multiple transmitters, including both a classical transmitter (such as DA) and a peptide transmitter (such as neurotensin). Indeed, it appears that few, if any, neurons contain only one transmitter, and in many if not most cases three or more transmitters are found in a single neuron.

The presence of multiple transmitters in single neurons suggests that different transmitters are used by a neuron to signal different functional states to its target cell. The firing rates of the neurons differ considerably, and it may be useful to encode fast firing by one transmitter and slower firing by a different transmitter. The firing *pattern* of neurons also conveys information. For example, a neuron may discharge five times every second on average; this may mean that the cell regularly

discharges every 200 ms, but can also represent a cell that has a burst of five discharges during an initial 100-ms period followed by 900 ms of silence. Peptide transmitters often are released at higher firing rates and particularly under burst-firing patterns.

The different synthetic steps in peptides and classical transmitters lead to differential release. Classical transmitters can be replaced rapidly because their synthesis occurs in nerve terminals, with rapid uptake of the released transmitter and energy-efficient recycling of the transmitter. In contrast, peptide transmitters must be synthesized in the cell body and transported to the terminal. Thus, it is useful to conserve peptide transmitters for situations of high demand because they would otherwise be depleted rapidly.

Transmitter Release from Different Processes

The restatement of Dale's principle by Eccles held that a transmitter is found in all processes of a neuron. In invertebrate species a transmitter can be localized to different parts of a neuron, and it is now clear that most proteins in mammalian neurons are not randomly distributed but present in distinct "compartments" of a neuron, often being reversibly tethered there by anchoring proteins. For example, if a transmitter were restricted to a particular part of a neuron, the neuron would need multiple transmitters to account for different release sites. Receptors are very specific in their locations on neurons, and certain receptors are recruited to "hot spots" on neuronal processes by the activity patterns of presynaptic inputs.

Synaptic Specializations versus Nonjunctional Appositions between Neurons

The anatomical relations between one cell and its follower may contribute to the need for different transmitters. We usually think of synaptic specializations as the physical substrate of communication between two neurons. However, nonsynaptic communication between two neurons can also take place, including over longer distances than that of conventional synapses. In such a situation the requirements for transmitter action would differ from those discussed previously because the distance traversed by the transmitter molecule would be farther than at a synaptic apposition. Thus, transmitters that lack an efficient reuptake system, such as peptide transmitters, might be favored at nonsynaptic sites. Because a single neuron can form both synaptic and nonsynaptic specializations, a single neuron may require more than one neurotransmitter.

Fast versus Slow Responses of Target Neurons to Neurotransmitters

Different firing rates or patterns of firing may be accompanied by changes in the type or relative amounts of a transmitter being released from a neuron. For example, stimulation of receptors that form ion channels leads to very rapid changes, whereas actions at metabotropic receptors that are coupled to intracellular events through specific transduction molecules have slower response characteristics. Differences in temporal response characteristics allow the receptive neuron to respond differently to a stimulus depending on the antecedent activity in the cell. A transmitter can change the response characteristics of a particular cell to subsequent stimuli by seconds or even minutes, and thus short-term changes can occur independent of changes in gene expression.

PEPTIDE TRANSMITTERS

Synthesis and Storage of Peptide Transmitters

Classical transmitters are synthesized in the axon from which they are released. In contrast, genes encoding most peptide transmitters give rise to a prohormone, which is incorporated into secretory granules, after which the prohormone is acted on by peptidases to form the peptide transmitter (Fig. 6.4). This process typically occurs in the cell body, and the peptide-containing vesicles are then transported to the axon; a small number of peptide transmitters are synthesized enzymatically.

In neurons that use classical transmitters, demands for increased amounts of transmitter are met by increasing local transmitter synthesis or activity at the axon terminal. However, increasing the amount of a peptide transmitter requires an increase in gene expression to yield a prohormone, with a subsequent delay in delivery to the axon terminal of the peptide. Thus, classical transmitters respond to increased demand rapidly, but peptide transmitters cannot.

The storage of peptides and classical transmitters also differs. Classical neurotransmitters are usually stored in small (~50 nm) synaptic vesicles, while neuropeptide transmitters are stored in large (~100 nm) dense-core vesicles. Because peptide transmitters are usually released at a high neuronal firing frequency or when neurons show a burst-firing pattern, different mechanisms may be involved in the release of peptide- and classical transmitter-containing vesicles. Although the release of both classical and peptide transmitters is calcium-dependent, distinct but related molecular mechanisms subserve the release of small and large dense-core vesicles (Sieburth, Madison, & Kaplan, 2007).

Inactivation of Peptide Transmitters

There are major differences in how peptide and classical transmitters are inactivated. Classical transmitter

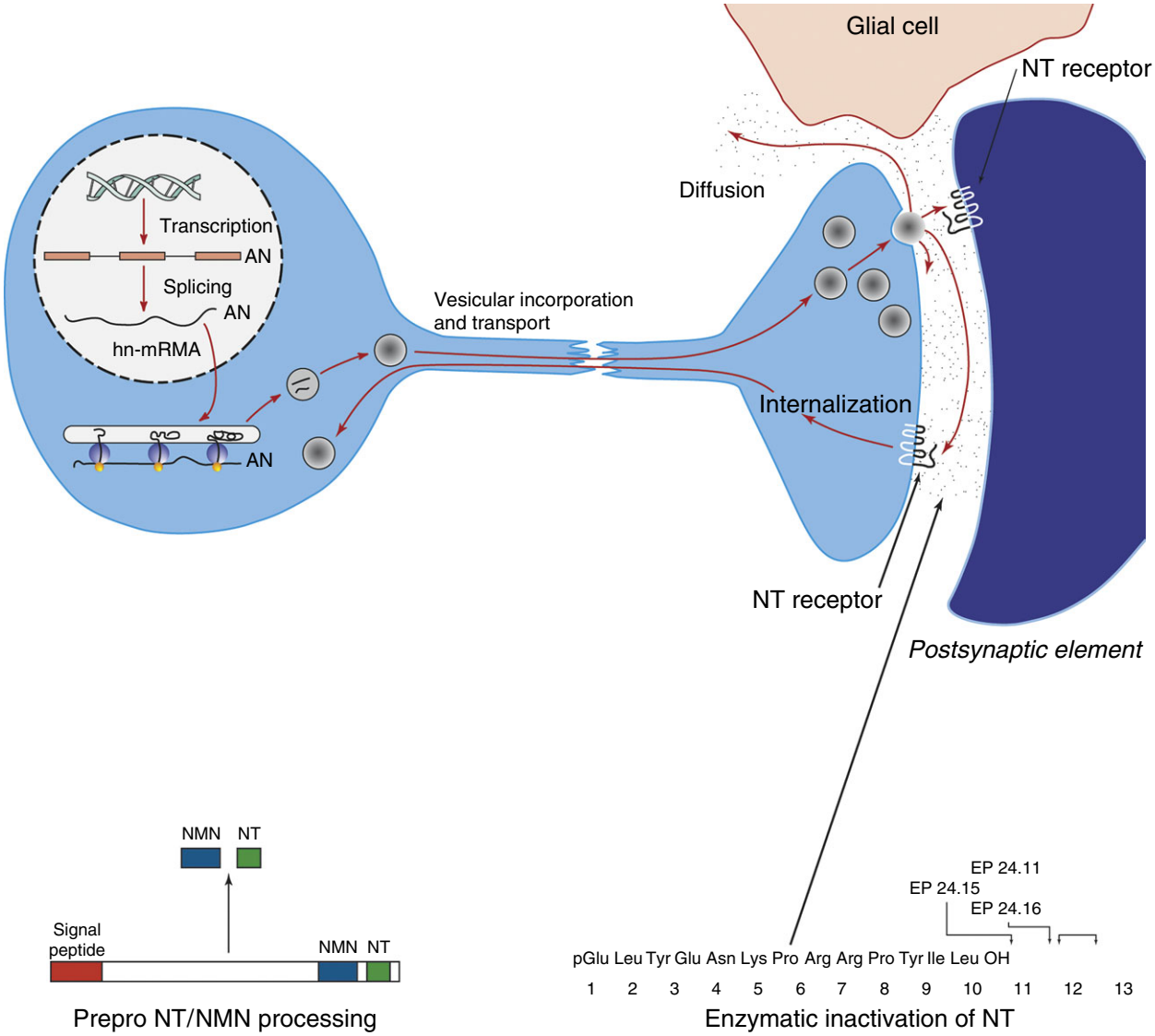


FIGURE 6.4 Schematic illustration of the synthesis, release, and termination of action of the peptide transmitter neurotensin (NT). The illustrative panels on the bottom show processing of NT from the prohormone (left) and enzymatic inactivation of NT (right).

actions primarily depend on a high-affinity reuptake process to remove the transmitter from the extracellular space. There are no such transporters for peptide transmitters, which are inactivated enzymatically or by diffusion.

Enzymatic inactivation of peptide transmitters differs from that of classical transmitters. Peptide transmitters are short chains of amino acids, but the inactivating enzymes are specific for certain dipeptide sequences and are not specific to any single peptide. For example, an enzyme that inactivates small opioid-like peptide transmitters (enkephalins) is often referred to as enkephalinase, but also cleaves other peptide transmitters.

A final difference between the inactivation of peptide and classical transmitters is the nature of the final product. When classical transmitters are broken down enzymatically, the metabolites are inactive at the transmitter's receptor. However, certain peptide fragments derived from the enzymatic "inactivation" of peptide transmitters are biologically active. An example is angiotensin, in which angiotensin I is metabolized to yield angiotensin II and III, each successively more active than the parent angiotensin I. It can therefore sometimes be difficult to distinguish between transmitter synthesis and transmitter inactivation. The peptide that is stored in vesicles and then released is therefore considered the transmitter, although the actions of certain peptidases may lead to other biologically active fragments. As a means of exploring aspects of the life cycle of peptide transmitters, we will now examine as a representative example the neuropeptide neurotensin.

Neurotensin Synthesis

Neurotensin (NT) is widely expressed in the central nervous system and in certain peripheral tissues, such as the small intestine. A related peptide, neuromedin N (NMN), is also present. A 170 amino acid prohormone precursor of NT is the product of a single gene that is transcribed to yield two mRNAs. The smaller transcript is the major form in the intestine; the two mRNA species are equally abundant in most brain areas. The precursor contains one copy each of NT and NMN. The molar ratios of NT/NMN vary across different tissues, suggesting different enzymatic processing of the prohormone or, alternatively, the generation of different transcripts. Because NT and NMN are contained in the same exon of the NT-NMN gene, differences in relative abundance of the two are due to differential processing of the precursor.

NT and other peptide transmitters are stored in large dense-core vesicles. Although there are some differences in the mechanisms controlling release of the contents of small and large dense-core vesicles, the general characteristics of NT release are similar to those of classical transmitters. Thus, depolarization evokes the Ca^{2+}-dependent release of both NT and NMN (Kitabgi, De Nadal, Rovere, & Bidard, 1992). The impulse-dependent release of NT varies as a function of frequency and pattern of impulses, with higher frequencies of firing or burst firing patterns eliciting greater peptide release (Bean & Roth, 1992).

Inactivation of Neurotensin

Peptides are inactivated by enzymatic actions or diffusion, or both. NT is degraded by three endopeptidases, grandly named 24.11, 24.15, and 24.16. Endopeptidase 24.11 cleaves NT at two specific sites to yield a decapeptide, as does endopeptidase 24.16 (Fig. 6.4); the other endopeptidase acts at a different set of amino acid residues. Because these enzymes act at dipeptides that are found in many peptides and proteins, and two of the three act at the same site, it is obvious that enzymes that inactivate peptides are not very specific.

Although there are no known membrane transporters for peptide transmitters, peptides can be accumulated by neurons. This occurs by internalization of a neuropeptide bound to its receptor. G protein-coupled receptors undergo internalization via an endocytotic mechanism, where they are either recycled to the membrane after various steps or shipped to lysosomes for degradation (Hanyaloglu & von Zastrow, 2008). However, once inside the cells the peptide and its receptor can dissociate, leaving the free neuropeptide to exert some actions.

Coexistence of Neurotensin and Classical Transmitters

Neurotensin is colocalized with dopamine in certain hypothalamic and midbrain neurons. The colocalization of NT and DA has provided a useful system in which to explore the interrelationships between two colocalized transmitters. In the prefrontal cortex of the rat, NT is found only in DA axons, although there are some DA axons that do not contain NT. Neurotensin release in the prefrontal cortex is increased when neuronal firing is increased or when DA neurons display a burst-like firing pattern (Bean & Roth, 1992). In addition, the release-modulating dopamine autoreceptor on DA axons regulates NT release, but in a very different manner than it regulates DA release: DA agonists (which decrease DA release) *enhance* NT release from colocalized DA-NT axons. Conversely, antagonists at the DA autoreceptor decrease NT release but enhance DA release. Thus, the release of NT and DA are reciprocally regulated by actions at release-modulating DA autoreceptors.

UNCONVENTIONAL TRANSMITTERS

The differences between peptide and classical transmitters substantially delayed the general acceptance of peptides as transmitters. However, once peptides were recognized as neurotransmitters, a gate opened for the consideration of radically different molecules as transmitters.

Remarkable technical advances have allowed us to measure substances in the brain that are present in minute quantities or are very unstable. The early "classical" neurotransmitters were present in relatively high concentrations, at least in some brain sites, permitting the measurement of these compounds by neurons after electrical or chemical stimulation. Peptide transmitters are usually found in smaller amounts, and sensitive biochemical assays for measuring neuropeptides were not widely used until the 1970s. When it became clear that there were many compounds, including certain peptides, that might potentially serve as chemical messengers between neurons, the basic criteria for designation of a molecule as a neurotransmitter came into question.

What should be the key attribute(s) of a neurotransmitter? One simple approach would be to designate as a neurotransmitter any compound that permits information flow from one neuron to another. However, this definition could encompass hormone signaling and does not address the temporal characteristics of transmitter action. Moreover, it does not address the possibility of a neurotransmitter being a molecule that conveys information between neurons and non-neuronal brain cells, such as glia. Finally, this definition does not accommodate unconventional roles for transmitters, such as the

regulation of neuronal development or intracellular trafficking of proteins. We will now discuss what we call, for lack of a better term, "unconventional" transmitters.

Endocannabinoids

The psychoactive properties of marijuana have been known for thousands of years. Marijuana contains dozens of structurally related cannabinoids. Cannabinol was the first of the cannabinoids to be structurally characterized. The major psychoactive cannabinoid, Δ-9-tetrahydrocannabinol (THC), was subsequently identified. The fact that marijuana and THC had psychoactive properties in humans led to search for a receptor for THC, the first of which, CB1, was cloned a full half century after the structure of cannabinol was first elucidated. As has been the case for other "orphan" receptors, the search for an endogenous ligand that binds to the CB1 receptor culminated two years after the cloning of CB1 with the identification of a compound that bound with high affinity to the receptor. Over the past 15 years, the identification of multiple endocannabinoids (ECs), the enzymes involved in synthesizing and degrading endocannabinoids, and the characterization of two cannabinoid receptors have led to the realization that ECs serve as chemical messengers between neurons.

Synthesis, Inactivation, and Release of Endocannabinoids

Endocannabinoids are hydrophobic lipids that pass easily through plasma membranes. This free diffusion across membranes means that ECs cannot be stored in vesicles. Instead, they are synthesized and released when needed from lipid precursors located in neuronal membranes. The basic mechanisms involved in the synthesis of ECs are reasonably well characterized, although some issues regarding regulation of EC synthesis have not been resolved (Alger & Kim, 2011).

There are two major ECs in brain, N-arachidonoylethanolamide (anadamide) and 2-arachidonoylglycerol (2-AG). The synthesis of the latter has been reasonably well characterized. An increase in intracellular calcium stimulates phospholipase C (PLC), which metabolizes membrane phosphoinositides to diacylglycerol (DAG). In turn, DAG is metabolized by DAG lipase to 2-AG. There is also a second minor route through which DAG can be produced from phosphatidic acid.

Anandamide is a fatty acid amide and is the best characterized of the ECs. However, the synthesis of anadamide can be accomplished through alternative pathways, and it is unclear which of these multiple synthetic routes predominates *in vivo*.

The membrane localization of the EC precursors for anandamide and 2-AG, coupled with the highly permeable nature of the two ECs, initially led to controversy concerning the ability of ECs to be synthesized only when the ("on-line") need for their release occurs. However, newer data favor the idea that ECs are indeed synthesized on demand.

Three proteins play prominent roles in EC inactivation. Two are enzymes (fatty acid amide hydrolyase [FAAH] and monoacylglycerol lipase [MGL]); the third protein is an anadamide transporter. 2-AG is metabolized primarily by MGL, but it can also be acted on by another enzyme (β hydrolase 6). In contrast, anandamide is metabolized by fatty acid hydrolase. The availability of mice with genetic deletions of EC synthetic or degradative enzymes has been critical in establishing the functional roles of anandamide and 2-AG. The anandamide transporter is a sodium- and energy-dependent carrier protein that facilitates diffusion.

Recent studies have demonstrated that ECs act as retrograde messengers at synapses. Thus, ECs are produced on demand in an activity-dependent manner by a neuron through the cleavage of membrane lipids. The resultant ECs travel backward (retrogradely) across the synapse to activate presynaptic cannabinoid receptors on axon terminals, resulting in a suppression of transmitter release. Because the presynaptic CB1 cannabinoid receptor is found on both GABAergic and glutamatergic nerve endings, EC signaling can modulate both excitatory and inhibitory transmission. EC-mediated retrograde signaling appears to be crucial for certain types of short- and long-term synaptic plasticity that underlie learning and memory.

Although still early, studies of ECs have inspired the development of novel treatment strategies for neuropsychiatric disorders. Consistent with subjective reports that marijuana increases eating (the "munchies"), early clinical trials reported some success of CB1 receptor antagonists in the treatment of obesity. However, other potential behavioral side effects of CB1 antagonists led to a halt in the development of these drugs, at least for the time being. Similarly, CB1 antagonists have been advanced as an aide to smoking cessation.

Gases as Unconventional Transmitters

Nitrates such as nitroglycerine have long been used to treat angina pectoris, in which pain arises because of insufficient blood delivery to the heart. Nitrates dilate cardiac blood vessels, thereby increasing blood flow and diminishing or relieving the pain. However, the mechanism by which nitrates dilate coronary arteries was not known until relatively recently. In 1980, a "relaxing" factor contained in the cells lining blood vessels was found to potently and rapidly dilate blood vessels. This factor was soon shown to be the gas nitric oxide (NO). Other studies found that the transmitter glutamate released a factor that caused blood vessels

in the brain to dilate. It then became apparent that endothelial-derived relaxing factor and the glutamate-induced factor that caused vasodilation were the same compound, NO. Thus, NO was present in neurons as well as cells forming blood vessels. These data led to the proposal that NO may take part in intercellular communication (see Box 6.2).

Nitric oxide is a well-known air pollutant. The idea that an unstable toxic gas could serve as a transmitter led to several incredulous but pointed questions, including "how can a gas be stored for release in an impulse-dependent manner?" The answer is simple: it is not stored. If one accepts the argument that NO is a transmitter, the classical definition of a neurotransmitter becomes untenable. Many theories can accommodate one exception. However, NO is not the only gaseous neurotransmitter. Other gases, including carbon monoxide and hydrogen sulfide, also play transmitter-like roles (Mustafa, Gadalla, & Snyder, 2009), leading to the introduction of the term gasotransmitter!

Synthesis of Nitric Oxide

Nitric oxide synthesis involves a single step, the conversion of L-arginine to NO and citrulline. The enzyme responsible for NO synthesis is aptly named nitric oxide synthase (NOS). There are three isoforms of NOS. Endothelial is found in the cells lining blood vessels, where the NO story began. Macrophage-inducible NOS is present in microglia, while the third isoform is localized to neurons (nNOS). All three isoforms require tetrahydrobiopterin as a cofactor and NADPH as a coenzyme. The presence of NADPH explains why the enzyme NADPH diaphorases, which had been described well before NOS was known, is found in high concentration in nNOS-positive neurons.

Storage and Release of Nitric Oxide

The subheading "Storage and Release of Nitric Oxide" promises a discussion of the storage of NO. As previously discussed, storage of a compound (usually in a place poised for release) was originally considered to be necessary for designation of a molecule as a transmitter. However, there is no storage of NO. Instead of NO being synthesized in a neuron and released in a calcium-dependent exocytosis from vesicles, NO is an uncharged molecule that diffuses freely across cellular membranes, including the membranes of vesicles.

The ability of NO to diffuse across cell membranes opens the possibility that NO serves as a chemical messenger to convey information regarding the activity of

BOX 6.2

GOING FOR GASES AS NEUROTRANSMITTER

One of the most rewarding experiences for a scientist is to find that long-held prejudices are altogether wrong and that a new, correct insight reveals a novel scientific principle. In the late 1950s, only acetylcholine and the biogenic amines were known to be neurotransmitters. The next decade saw amino acids acknowledged as neurotransmitters. The discovery of enkephalins and endorphins in the mid-1970s reinforced gradually accumulating evidence that peptides are transmitters, and now we find that there are over 100 different bioactive brain peptides.

In the 1970s and early 1980s, nitric oxide (NO) was found to mediate the ability of macrophages to kill tumor cells and bacteria and to regulate blood vessel relaxation. A short report suggested that NO can be formed in brain tissue. Progress in the NO field at that time was slow because assaying NO synthase (NOS), the enzyme that oxidizes the amino acid arginine to NO, was quite tedious, based on the accumulation of nitrite formed from the NO. A much simpler approach was to monitor the conversion of [^{3}H]arginine into [^{3}H]citrulline, which is formed simultaneously with NO; this assay could process 100 or more samples in an hour. Research on blood vessels revealed that NO acts by stimulating cyclic GMP formation. In the brain, the excitatory neurotransmitter glutamate was known to augment cyclic GMP levels. Glutamate, acting through its NMDA receptor, triples NO synthase activity in a matter of seconds, and arginine derivatives that inhibit NOS activity block the elevation of cyclic GMP. This finding causally linked the actions of so prominent a neurotransmitter as glutamate to NO.

To determine if NO was a neurotransmitter, it was necessary to ascertain whether NOS was localized in neurons. The most straightforward approach would be to generate an antibody to use in anatomical studies. However, purifying NOS protein to generate an antibody proved very difficult because the enzyme lost its activity in attempts to purify it. The addition of calmodulin was found to stabilize the enzyme. Because calmodulin is a calcium-binding protein, this finding immediately explained how NO formation can be triggered rapidly by synaptic activation through glutamate. When glutamate activates its NMDA receptor, Ca^{2+} rushes into the cell, binds to calmodulin, and activates NOS.

The ability of purified NOS led to antibodies being developed and NOS being localized by immunohistochemistry. The neuronal form of NOS (nNOS) is present

BOX 6.2 (*cont'd*)

in only about 1% of the neurons in the brain. However, these cells give rise to processes that ramify so extensively that probably every neuron in the brain is exposed to NO. The purified NOS protein also allowed an amino acid sequence to be obtained and the gene for the enzyme cloned. The structure of NOS revealed that it is regulated by many more factors than virtually any other enzyme in biology, including at least five oxidative-reductive cofactors, four phosphorylating enzymes, and three binding proteins. This makes sense because of the unique properties of NO as a gaseous neurotransmitter. Most neurotransmitters are stored in vesicles with large storage pools so that only a small amount of the transmitter is released with each nerve impulse. In contrast, every time a neuron wishes to release a molecule of NO, it must activate NOS—hence, a requirement for exquisitely subtle regulation of the enzyme.

Neurotransmitters come in chemical classes such as biogenic amines, amino acids, and peptides. Might there be at least one other gaseous neurotransmitter? Carbon monoxide (CO) is normally formed in the body by the enzyme heme oxigenase, which is primarily responsible for degrading heme in aging red blood cells. It cleaves the heme ring to form biliverdin, which is reduced rapidly to bilirubin, the pigment that accounts for jaundice in patients with a degradation of red blood cells. When the enzyme cleaves the heme ring, CO is released as a single carbon fragment.

The biosynthetic enzyme for a transmitter should be localized to selected neuronal populations. Heme oxigenise-2, the neuronal form of the enzyme, was shown to be localized to discrete neuronal populations throughout the brain. To seek a neurotransmitter function, the peripheral nervous system (in which synaptic transmission is characterized more readily than in the brain) was used. The myenteric plexus of nerves regulates intestinal peristalsis. A previously unidentified neurotransmitter of myenteric plexus neurons accounts for the relaxation phase of peristalsis. nNOS had already been localized to neurons of the myenteric plexus, and some functional evidence showed that NO might be a neurotransmitter of this pathway. Heme oxigenise-2 was found to be localized to the same myenteric plexus neurons as NOS. Mice with targeted deletions of the genes for nNOS or heme oxigenise-2 were used to elucidate function. In both types of gene knockout mice, intestinal relaxation evoked by neuronal depolarization was reduced about 50%, implying that NO and CO each contribute half of the relaxation. This finding, along with other evidence, established transmitter functions for both NO and CO and suggested that they are functioning as cotransmitters, although exactly how they interact remains a mystery. Such a cotransmitter role reminds us of the fact that most neurons in the brain contain at least two and sometimes more neurotransmitters. Thus, in addition to overturning a number of dogmas about neurotransmission, NO and CO may help resolve the riddle of cotransmission.

Solomon H. Snyder

one cell to a number of other cells, which can be pre- or postsynaptic to the neuron from which NO is liberated, or even be non-neuronal cells (in which case it is not synaptic transmission in the original sense of the phrase). Nitric oxide release from neurons occurs in response to glutamate release and subsequent activation of N-methyl-D-aspartate (NMDA) receptors (and voltage-dependent calcium channels) on the postsynaptic cell that allows calcium to enter the neuron. The increase in intracellular calcium in turn activates the enzyme nNOS, resulting in diffusion of NO to reach nearby cells.

Termination of Action of Nitric Oxide

The formation and diffusion of NO mean that it can affect many nearby cells. Although no receptors for NO have been identified, cell surface receptors are not necessary for a molecule that freely diffuses; once inside of nearby neurons and other cells NO can act on second messenger cascades (notably cGMP-mediated events; see Chapter 8).

We previously noted that unrestrained release of a neurotransmitter does not convey to the receptive follower cells information regarding the activity of a presynaptic neuron and is often deleterious. Because NO is a gas, one might envision a molecule of NO careening through neurons like a crazed arrow. There are no degradative enzymes that inactivate NO, nor are there transporters that accumulate NO that is in the extracellular space for a fleeting time. Instead, the inactivation of NO is passive, with NO spontaneously decaying to a nitrate, with a half life of about 30 seconds.

Actions of NO on Neurons

Once in the cytosol of neurons and other cells in the brain, NO can exert function effects by stimulating guanylyl cyclase to produce cGMP levels and initiating a cascade of intracellular effectors. Because NO is produced in response to activation of nNOS by a glutamate-mediated increase in intracellular calcium levels, NO is mainly synthesized in the dendrites of neurons. The

diffusion of NO starts from the dendrites and then moves unpredictably to nearby cells, many of which are presynaptic to the released glutamate that started NO's brief wandering. We have previously discussed a number of mechanisms that inform the presynaptic neuron of its activity, allowing for feedback modulation; these include processes ranging from release-modulating autoreceptors to end-product inhibition of biosynthetic enzymes. Because NO diffuses freely and can influence neurons both pre- and postsynaptic to the cell from which it originates, NO (like endocannabinoids) can provide a "report" to the presynaptic neuron about the physiological effect NO is having on the postsynaptic cell. In this way, "synaptic transmission" becomes bidirectional.

Nitric Oxide as a Neurotransmitter

The list of exceptions posed by NO to the dogma of traditional neurotransmitters is long. NO is not stored in cells, is not released in an exocytotic manner, lacks an active process that terminates its action, does not interact with specific membrane receptors on target cells, and often acts as a retrograde transmitter to regulate axon terminals presynaptic to the neuron in which NO is synthesized. It is not difficult to understand the skepticism that first met the hypothesis that NO is a neurotransmitter, or to have some sympathy for those who expressed the view that NO is not a transmitter but an alien event, benign or otherwise, intent on making neuroscientists question their most cherished beliefs (Fig. 6.5).

Summary

We have discussed two "unconventional" transmitters, one a gas and the other a lipid. This growing category of transmitters contains many other similarly unconventional molecules, including growth factors, neuroactive steroids, and others still to be uncovered. All offer a significant challenge to our definitions of what constitutes a neurotransmitter, but with the advantage of expanding our knowledge of how the nervous system works in health and disease.

THE EXPANDING SYNAPSE AND GLIOTRANSMISSION

Our discussion of neurotransmission has been focused on the molecular signaling between pre- and postsynaptic neurons necessitated by the presence of a synaptic cleft that separates the two neurons. However, there are cells in the brain other than neurons, including different types of glial cells. The processes of one of these glial cells, the astrocyte, have long been known to be closely positioned near synapses and to be closely apposed to blood vessels. The latter structural arrangement was thought to contribute to, in part, the blood–brain barrier. The astrocyte was also known to play important roles in potassium homeostasis in the brain, as well as in scavenging cellular debris after brain injury.

However, as noted in earlier parts of this chapter, various molecules involved in the life cycle of classical

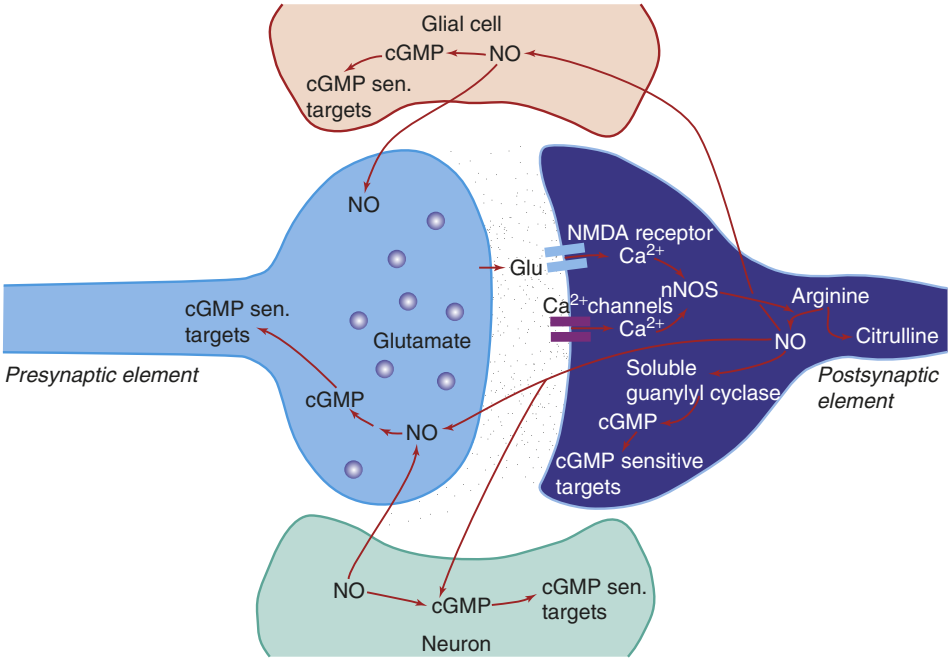


FIGURE 6.5 Schematic representation of a nitric oxide (NO)-containing neuron. NO is formed from arginine by the actions of different nitric oxide synthases (NOS). NO diffuses freely across cell membranes and can thereby influence both presynaptic neurons (such as the glutamatergic presynaptic neuron in the figure) or other cells that are not apposed directly to the NOS-containing neuron; these other cells can be neurons or glia.

transmitters, particularly glutamate and GABA, are expressed not only by neurons but also by astrocytes. For example, the major glutamate transporter involved in termination of action of glutamate is expressed by astrocytes rather than neurons. A metabotropic glutamate receptor that functions as a glutamate autoreceptor is also expressed by astrocytes, as is the enzyme responsible for the synthesis of kynurenic acid. These and similar observations raised the possibility that astrocytes play an important role in synaptic transmission, and coupled with structural data led to the formulation of the "tripartite synapse" (Araque, Parpura, Sanzgiri, & Haydon, 1999) involving classic pre- and postsynaptic neuronal elements and astrocytes.

The past decade has seen a dramatic rise in our understanding of the multiple roles played by astrocytes (see Allen & Barres, 2009; Halassa & Haydon, 2010; Panatier et al., 2011). Astrocytes repond to neuronal activity by increases in intracellular calcium, which can lead to the release of several transmitters from astrocytes; the precise mechanisms of the release of these gliotransmitters are not completely understood. Among these transmitters are glutamate, D-serine (which binds to a site on the glutamatergic NMDA receptor complex), kynurenic acid, and ATP. Because astrocytic processes of a single astrocyte can envelope thousands of synapses, gliotransmitters can coordinately regulate a very large number of neurons. In addition, the astrocytic "end feet" apposed to capillaries, which were thought to be a structural compartment of the blood–brain barrier, are now thought to be involved in glucose transport from the microvasculature to the brain and thus be critical to the metabolism and activity of neurons. Finally, glia have recently been shown to be critically involved in synaptogenesis and the pruning of synapses seen in the brain during development (Eroglu & Barres, 2010).

These transmitter-like roles played by astrocytes have recently been explored in the context of neuropsychiatric disorders, including substance abuse, schizophrenia, depression, and sleep disorders. Thus, in a very short period of time, glia have been shown to be critical partners with neurons and be involved in the pathophysiology of a large (and growing) number of disorders.

Concluding Remarks

Several of the key proteins involved in different aspects of chemical neurotransmission in mammals were originally identified on the basis of homologies to proteins found in invertebrates, such as the worm *C. elegans* or the fly *Drosophila melanogaster*. Some of the same molecules used as neurotransmitters in animals can be found in plants, along with their receptors. Although we tend to think of neurotransmitters, their synthesizing and degradative enzymes, and their transporters as having a single or very restricted roles, as nervous systems have evolved, many transmitter-related proteins have maintained roles that are not directly related to transmitter function or, alternatively are involved in less discrete and more spatially elaborate signaling.

A good example of this is acetylcholine, the synthesis of which depends on the enzyme choline acetyltransferase. The mRNA for ChAT is not only found in the brain, but is also present in the testes, where it is functional and contributes to the appearance of acetylcholine in spermatozoa. ChAT mRNA has also been reported to be present in lymphocytes, as have certain muscarinic cholinergic receptors. Also found in these white blood cells is the mRNA encoding AChE. Decreases in both acetylcholinesterase and butyrylcholinesterase enzyme activity have been reported in lymphocytes in Alzheimer's disease. Still another example is the presence of AChE in bone marrow and peripheral blood cells in certain leukemias; recent data indicate that inhibition of AChE gene expression in bone marrow suppresses programmed cell death (apoptosis; see Chapter 17). These examples clearly indicate that although a neurotransmitter may have one role in the brain, the same molecule may play a totally different role in other tissues. Such multifunctional roles of transmitters and related proteins should at the very least encourage us to attend to developments in diverse fields.

We have discussed chemically coded synaptic transmission. However, one cannot discuss the biochemistry and pharmacology of synaptic transmission without referring to and appreciating critical information about the structure (anatomy) and function (physiology and behavior) of neurons. Neuroscience is multidisciplinary, requiring an understanding of different aspects of cellular function to come to grips with the basic principles of synaptic communication.

Our concepts of synaptic transmission are in a greater degree of flux than ever before, and require frequent reevaluation and revision. Although some aspects of neurotransmission have been known for about a 100 years, others have become apparent only within the past year. Classical transmitters are but one branch of the family of transmitters, with other relatives including peptides, gases, growth factors, and lipid-derived transmitters such as endocannabinoids. The synapse has historically been viewed as the close apposition of two neurons. It is now clear that astrocytes, once viewed as structural support elements of the brain, play critical roles in transmitter function, ranging from terminating the actions of glutamate to promoting organized multineuronal activation patterns. The use of the terms "conventional" and "unconventional" in discussing transmitters is indicative of our

unease with the expanding definition of neurotransmitters. Because our knowledge of neuroscience is constantly changing we must be open and receptive to challenges to long-taught dogma.

References

Agnati, L. F., Guidolin, D., Guescini, M., Genedani, S., & Fuxe, K. (2010). Understanding wiring and volume transmission. *Brain Research Reviews, 64,* 137–159.

Alger, B. E., & Kim, J. (2011). Supply and demand for endocannabinoids. *Trends in Neurosciences, 34,* 304–315.

Allen, J. A., & Barres, B. A. (2009). Glia–more than just brain glue. *Nature, 457,* 675–677.

Araque, A., Parpura, V., Sanzgiri, R. P., & Haydon, P. G. (1999). Tripartite synapses: Glia, the unacknowledged partner. *Trends in Neurosciences, 22,* 208–215.

Bean, A. J., & Roth, R. H. (1992). Dopamine-neurotensin interactions in mesocortical neurons: Evidence from microdialysis studies. *Annals of the New York Academy of Sciences, 668,* 43–53.

Cartmell, J., & Schoepp, D. D. (2000). Regulation of neurotransmitter release by metabotropic glutamate receptors. *Journal of Neurochemistry, 75,* 889–907.

Chaudhry, F. A., Lehre, K. P., van Lookeren Campagne, M., Otterson, O. P., Danbolt, N. C., & Storm-Mathisen, J. (1996). Glutamate transporters in glial plasma membranes: Highly differentiated localizations revealed by quantitative ultrastructural immunocytochemistry. *Neuron, 15,* 711–720.

El Mestikawy, S., Wallén-Mackenzie, A., Fortin, G. M., Descarries, L., & Trudeau, L. E. (2011). From glutamate co-release to vesicular synergy: Vesicular glutamate transporters. *Nature Reviews Neuroscience, 12,* 204–216.

Erlander, M. G., & Tobin, A. J. (1991). The structural and functional heterogeneity of glutamic acid decarboxylase: A review. *Neurochemical Research, 16,* 215–226.

Eroglu, C., & Barres, B. A. (2010). Regulation of synaptic connectivity by glia. *Nature, 468,* 223–231.

Fremeau, R. T., Voglmaier, S., Seal, R. P., & Edwards, R. H. (2004). VGLUTs define subsets of excitatory neurons and suggest novel roles for glutamate. *TINS, 27,* 98–102.

Gainetdinov, R. R., Jones, S. R., Fumagalli, E., Wightman, R. M., & Caron, M. G. (1998). Re-evaluation of the role of the dopamine transporter in dopamine system homeostasis. *Brain Research Reviews, 26,* 148–153.

Greenfield, S. A., Zimmermann, M., & Bond, C. E. (2008). Non-hydrolytic functions of acetylcholinesterase: The significance of C-terminal peptides. *FEBS Journal, 275,* 604–611.

Halassa, M. M., & Haydon, P. G. (2010). Integrated brain circuits: Astrocytic networks modulate neuronal activity and behavior. *Annual Review of Physiology, 72,* 335–355.

Hanyaloglu, A. C., & von Zastrow, M. (2008). Regulation of GPCRs by endocytic membrane trafficking and its potential implications. *Annual Review of Pharmacology and Toxicology, 48,* 537–568.

Hassel, B., & Dingledine, R. (2006). Glutamate. In Siegel, G. J., Albers, R. W., Brady, S. & Price, D. L. (Eds.), *Basic neurochemistry*. (7th ed., pp. 267–290). San Diego, CA: Elsevier–Academic Press.

Iversen, L., Iversen, S., Bloom, F. E., & Roth, R. H. (2008). *Introduction to neuropsychopharmacology*. New York, NY: Oxford University Press.

Jiang, J., & Amara, S. G. (2011). New views of glutamate transporter structure and function: Advances and challenges. *Neuropharmacology, 60,* 172–181.

Kitabgi, P., De Nadal, F., Rovere, C., & Bidard, J. N. (1992). Biosynthesis, maturation, release, and degradation of neurotensin and neuromedin N. *Annals of the New York Academy of Sciences, 668,* 30–42.

Kohara, K., Kitamura, A., Morishima, M., & Tsumoto, T. (2001). Activity-dependent transfer of brain-derived neurotrophic factor to postsynaptic neurons. *Science, 291,* 2419–2423.

Mustafa, A. K., Gadalla, M. M., & Snyder, S. H. (2009). Signaling by gasotransmitters. *Science Signaling, 2,* . re2.

Olsen, R. W., & Betz, H. (2006). GABA and glycine. In Siegel, G. J., Albers, R. W., Brady, S. & Price, D. L. (Eds.), *Basic neurochemistry*. (7th ed., pp. 291–302). San Diego, CA: Elsevier–Academic Press.

Panatier, A., Vallée, J., Haber, M., Murai, K. K., Lacaille, J. C., & Robitaille, R. (2011). Astrocytes are endogenous regulators of basal transmission at central synapses. *Cell, 146,* 785–798.

Schwarcz, R. (2004). The kynurenine pathway of tryptophan degradation as a drug target. *Current Opinion in Pharmacology, 4,* 12–17.

Shepherd, G. M. (1991). *Foundations of the neuron doctrine*. New York, NY: Oxford University Press.

Sieburth, D., Madison, J. M., & Kaplan, J. M. (2007). PKC-1 regulates secretion of neuropeptides. *Nature Neuroscience, 10,* 49–57.

Taylor, P., & Brown, J. H. (2006). Acetylcholine. In Siegel, G. J., Albers, R. W., Brady, S. & Price, D. L. (Eds.), *Basic neurochemistry*. (7th ed.). San Diego: Elsevier-Academic Press.

Valenstein, E. S. (2005). *The war of the soups and the sparks*. New York, NY: Columbia University Press.

Xu, E., Gainetdinov, R. R., Wetsel, W. C., Jones, S. R., Bohn, L. M., Miller, G. W., … Caron, M. G. (2000). Mice lacking the norepinephrine transporter are supersensitive to psycho-stimulants. *Nature Neuroscience, 3,* 465–471.

Zucchi, R., Chiellini, G., Scanlan, T. S., & Grandy, D. K. (2006). Trace amine-associated receptors and their ligands. *British Journal of Pharmacology, 149,* 967–978.

Ariel Y. Deutch

CHAPTER

7

Release of Neurotransmitters

The synapse is the primary place at which information is transmitted from neuron to neuron or from neuron to peripheral target, be it a gland or a muscle. Most synapses rely on a chemical intermediary, or transmitter, secreted in response to an action potential in the presynaptic cell in order to influence the activity of the postsynaptic cell. In chemical transmission, a single action potential in a small presynaptic terminal can generate a large *postsynaptic potential* (PSP) (as large as tens of millivolts). This is accomplished by the release of thousands to hundreds of thousands of molecules of transmitter that can bind to postsynaptic receptor molecules and open (or close) thousands of ion channels in about 1 ms. The effect can be either excitatory or inhibitory, depending on the ions that permeate the channels operated by the receptor. The resulting responses are either *excitatory postsynaptic potentials* (EPSPs) or *inhibitory postsynaptic potentials* (IPSPs), depending on whether they drive the cell toward a point above or below its firing threshold, as discussed in Chapter 10.

Why are most synapses chemical? The simpler alternative might appear to be the electrical synapse (see Chapter 9) in which the electrical signal from one cell can cross directly into the next through the electrically conducting pathway of the gap junction. Even when pre- and postsynaptic elements are coupled in this manner, however, a presynaptic spike of 100 mV would likely cause only a 1-mV change in the postsynaptic cell because relatively little current can flow through these junctions to charge the large membrane capacitance of the postsynaptic cell. More efficient electrical transfer can be had if the connection between the cells is large; the presynaptic terminal needs to be as large as or larger than the postsynaptic element and therefore only a few presynaptic cells could converge on a given postsynaptic cell. Moreover, such synapses could only induce excitation, not inhibition, in response to a presynaptic action potential and the computational capacity of a brain operated with just electrical synapses would be very low. Chemical synapses are characterized by great flexibility. Different afferents can have different effects, with different strengths and time courses, on each other as well as on postsynaptic cells. These differences depend on the identity of the transmitter(s) released and the receptors present (see Chapters 6 and 8). Chemical synapses often are modified by prior activity in the presynaptic neuron. Chemical synapses are also particularly subject to the modulation of presynaptic ion channels by substances released by the postsynaptic or neighboring neurons. This flexibility is essential for the complex processing of information that neural circuits must accomplish, and it provides an important locus for modifying neural circuits in adaptive processes such as learning (Chapters 47 and 48).

TRANSMITTER RELEASE IS QUANTAL

In order to secrete thousands of transmitter molecules rapidly and simultaneously, nerve terminals, like other secretory cells, package the transmitter into membrane-enclosed organelles. When one of these *synaptic vesicles* fuses with the plasma membrane (a process called *exocytosis*), the contents of the vesicle (approximately 5000 molecules) can diffuse into the extracellular space and encounter receptors on the postsynaptic cell within hundreds of microseconds. A consequence of releasing transmitter by the exocytosis of vesicles is that synaptic transmission is *quantal*: responses are built from signals with a discrete amplitude corresponding to a single vesicle. In the absence of presynaptic electrical activity, transmitter is released spontaneously as individual quanta. Each packet generates a small postsynaptic signal—either a *miniature excitatory* or a *miniature inhibitory postsynaptic potential* (MEPSP or MIPSP, respectively,

© 2013 Elsevier Inc. All rights reserved.

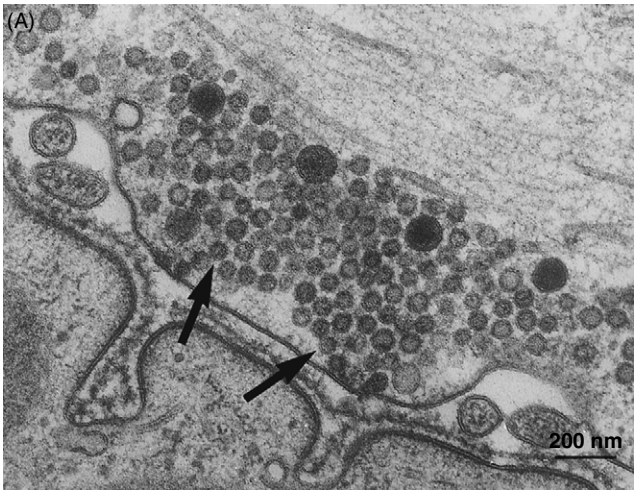
or just "mini")—that can be detected by microelectrode recording (Katz, 1969). An action potential accelerates tremendously, but very briefly, the rate of secretion of quanta and synchronizes them to evoke a PSP. At a synapse between two neurons, this might represent the release of 1 to 10 vesicles. At the vertebrate neuromuscular junction, a remarkably large and specialized synapse, hundreds of vesicles can be released, and the response is, to a first approximation, the sum of the individual quanta.

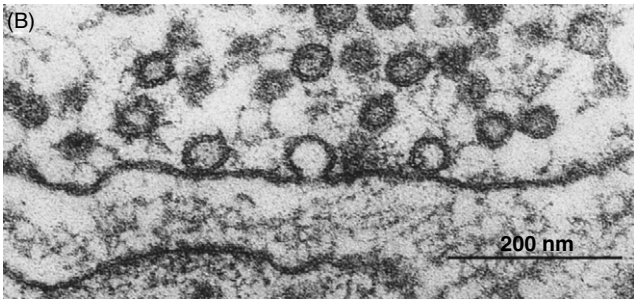
The anatomical specializations of the synapse and the properties of the presynaptic ion channels and postsynaptic receptors unite to achieve fast, quantal transmission (Figs. 7.1 and 7.2). The chief anatomical feature of the terminal is the profusion of synaptic vesicles, typically 50 nm in diameter, that cluster near the synapse and dock at specialized sites called *active zones* along the presynaptic membrane (Fig. 7.1). The vesicles may differ in their appearance depending on the transmitter they enclose: thus glutamate and acetylcholine are stored in small clear vesicles, whereas peptide neurotransmitters occupy large dense-cored vesicles. Although it may appear as indistinct fuzz on electron micrographs, the active zone is likely to be a highly structured specialization of the membrane and cytoskeleton (Fig. 7.1; also, see discussion later). Action potentials release transmitter by triggering the exocytosis of vesicles (Fig. 7.1B and 7.1C) and the release of their contents into the narrow *synaptic cleft* (about 100 nm wide) separating the presynaptic terminal from high concentrations of receptors on the postsynaptic membrane. At neuromuscular junctions, one of the best studied synapses, transmitter from one vesicle diffuses across the synaptic cleft in 2 ms and reaches a concentration of about 1 mM at the postsynaptic receptors. A large number of these receptors (up to 2000) will bind transmitter rapidly. Activation of a receptor entails the opening of an ion channel (see Chapter 8). Each channel has a 25 pS conductance and remains open for about 1.5 ms, admitting a net inflow of 35,000 positive ions. Thus, a single vesicle by opening 2000 channels can cause 70 million ions to cross the membrane and give rise to a mini of a few millivolts, the individual quantal response. The neuromuscular junction also is specialized for efficiency; a single nerve ending will have about 1000 active zones and a single action potential will cause a vesicle to fuse at about a third of those. The consequence is the release of about 300 quanta within 1.5 ms. The resulting postsynaptic depolarization begins after a synaptic delay of about 0.5 ms, reaches a peak of tens of millivolts, and is typically sufficient to generate an action potential in the muscle fiber.

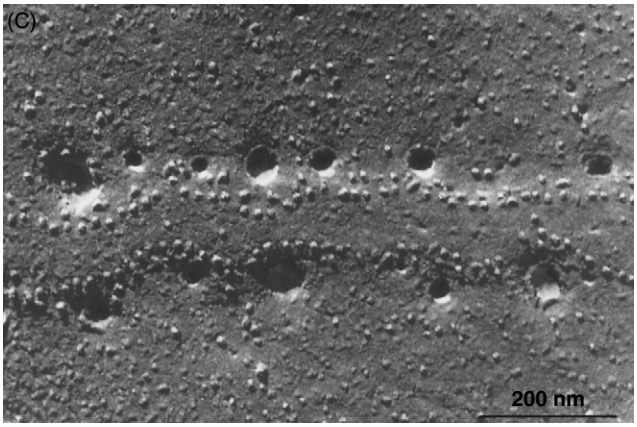
The anatomy and physiology are somewhat different at fast central synapses. Postsynaptic cells make contact with presynaptic axon swellings that are called *varicosities* when they occur along fine axons and are called *boutons* when they are located at the tips of axons. The postsynaptic cell can be contacted on the cell body, but often the synapse is made onto a fine dendritic branch or tiny spine with a length of a few micrometers. A typical varicosity or bouton contains one to four active zones. At any single active zone, an action potential may release zero, one, or perhaps two vesicles. However, with multiple active zones between two cells, the action of each zone will be additive in determining the response of the postsynaptic cell. At a representative excitatory glutamatergic synapse, each action potential releases from 5 to 10 vesicles. Each vesicle that is released elevates transmitter concentration in the cleft to about 1 mM, as at the neuromuscular junction, but because there are fewer receptors clustered beneath these synapses, each released vesicle activates only 30 or so ion channels as compared with the 1000–2000 channels activated at the neuromuscular junction. At excitatory synapses, the release of a vesicle may be sufficient to generate EPSPs of 1 mV or less in amplitude, clearly subthreshold for generating action potentials. However, central neurons often receive thousands of inputs, each of which has a "vote" on how the cell should respond.

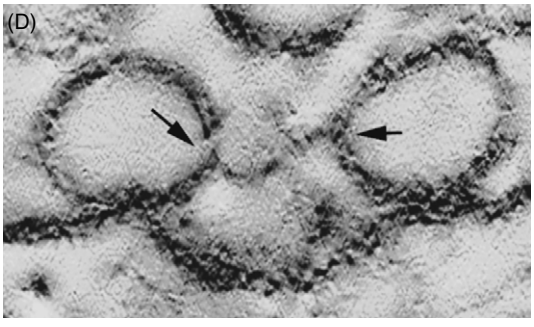
A comparison of these excitatory synapses onto small neurons with the vertebrate neuromuscular junction at the much larger skeletal muscle fiber illustrates two important points: (1) the fundamental quantal nature of the chemical synapse is universal and (2) the different physiological requirements of particular synapses, such as synapses onto small or large cells, can be met by

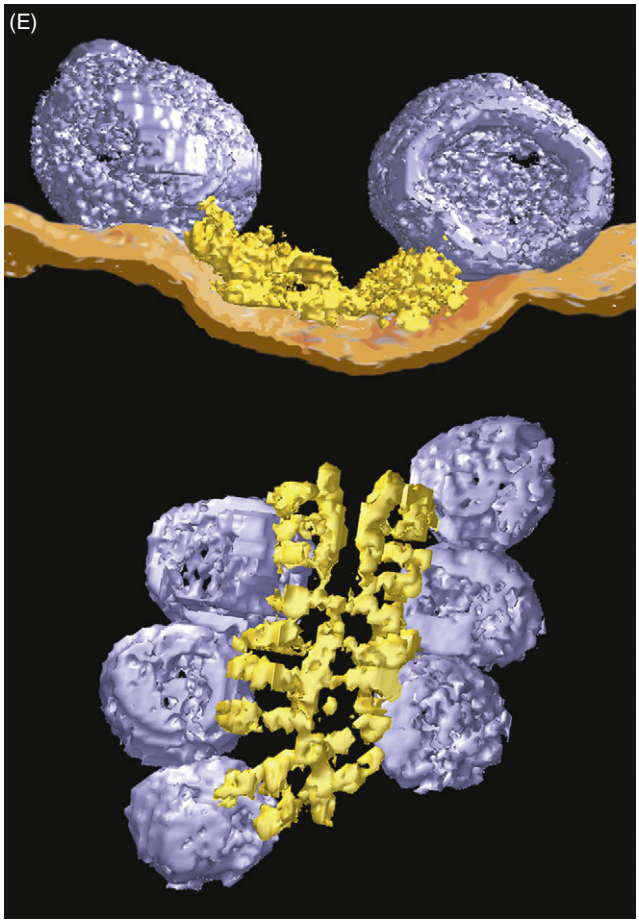
FIGURE 7.1 Ultrastructural images of exocytosis and active zones. (A–C) Synapses from frog sartorius neuromuscular junctions were quick-frozen milliseconds after stimulation in conditions that enhance transmission. (A) A thin section showing vesicles clustered in the active zone (arrows), some docked at the membrane. (B) Shortly (5 ms) after stimulation, vesicles were seen to fuse with the plasma membrane. (C) After freezing, presynaptic membranes were freeze-fractured and a platinum replica was made of the external face of the cytoplasmic membrane leaflet. Vesicles fuse about 50 nm from rows of intramembranous particles thought to include Ca^{2+} channels. (D–F) The fine structure of the active zone at a frog neuromuscular junction as seen with electron tomography. (D) In a cross-sectional image from tomographic data, two vesicles are docked at the plasma membrane and additional electron-dense elements are seen. When these structures are traced and reconstructed through the volume of the EM section (E), proteins of the active zone (gold) appear to form a regular structure adjacent to the membrane that connects the synaptic vesicles (silver) and plasma membrane (white). Viewed from the cytoplasmic side (E, lower image), proteins are seen to extend from the vesicles and connect in the center. (F) Schematic rendering of an active zone based on tomographic analysis. An ordered structure aligns the vesicles and connects them to the plasma membrane and to one another. *Parts A and B from Heuser (1977) and Heuser et al. (1979). Part C reproduced from the* Journal of Cell Biology ***88****, 564–580 (1981). (D–F) After Harlow et al. (2001).*

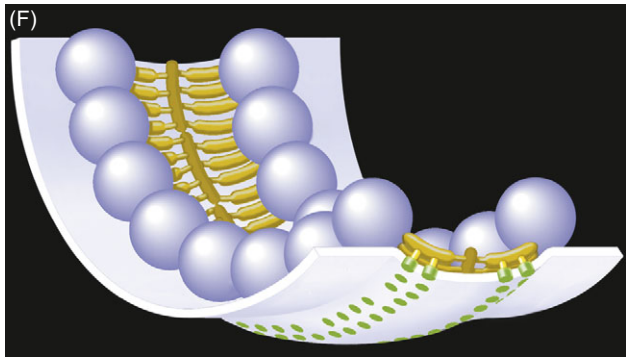

(A)
200 nm


(B)
200 nm


(C)
200 nm


(D)


(E)


(F)

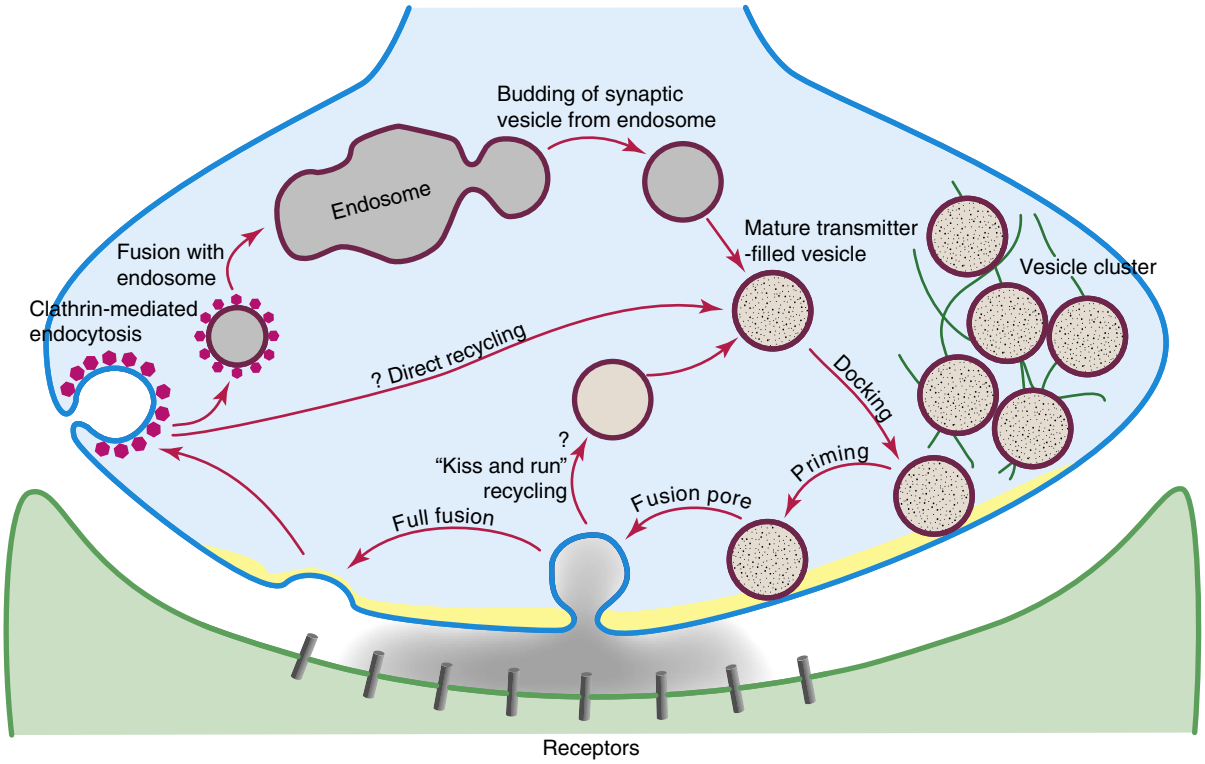


FIGURE 7.2 The life cycle of synaptic vesicles. Transmitter-filled vesicles can be observed in clusters in the vicinity of the active zone. Some vesicles are recruited to sites within the active zone in a process called docking. These vesicles subsequently are primed for release. The rise in cytosolic Ca^{2+} that occurs during an action potential triggers the opening of a fusion pore between some of the primed, docked vesicles and the plasma membrane. Transmitter exits through this fusion pore. Three pathways are proposed by which the now empty vesicle can be recovered and returned to the releasable pool: (1) by a direct reclosing of the fusion pore and reformation of the vesicle, often called "kiss and run"; (2) by complete fusion (i.e., the flattening of the vesicle onto the membrane surface) followed by clathrin-mediated endocytosis, removal of the clathrin coat, and return of the vesicle to the releasable pool; and (3) by complete fusion and recycling as in the second pathway, but the endocytosed vesicle fuses first with an endosome and mature vesicles are subsequently formed by budding from the endosome. After or during this recycling process, the vesicle must be refilled with transmitter.

differences in any of several parameters, including the number of receptors activated per vesicle released, the number of active zones connecting two cells, the probability of a given active zone releasing a vesicle, and the conductance of the postsynaptic receptor channel. In describing synaptic transmission, two parameters are particularly important. *Quantal size* or quantal amplitude describes the unitary response to release of a single quantum (vesicle), whereas *quantal content* refers to the average number of quanta (vesicles) released by a single impulse. These parameters are discussed further later.

EXCITATION–SECRETION COUPLING

The action potential is an electrical event—a change in the voltage gradient across the plasma membrane. How is this electrical change converted to the fusion of synaptic vesicles and the release of neurotransmitter? The classic studies of Bernard Katz and colleagues (Katz, 1969) established that this coupling is achieved by the use of Ca^{2+} as an intracellular messenger. Ca^{2+} inside the cell normally is buffered to very low levels (~100 nM) although the extracellular Ca^{2+} concentration will be in the millimolar range. Thus both concentration and electrical gradients provide a strong driving force for Ca^{2+} entry and, when a voltage-dependent Ca^{2+} channel opens in response to the depolarization of the membrane during an action potential, there is the potential for the intracellular Ca^{2+} concentration to increase 1000-fold. In this manner, the electrical signal is converted to a large chemical signal, and Ca^{2+} sensors in the fusion machinery can trigger vesicle fusion. The mechanism of fusion itself is discussed later, but the nature of the Ca^{2+} signal and the evidence for its centrality must be understood first.

It has been known for over 100 years that Ca^{2+} must be present in the extracellular saline for transmission to occur. Subsequently (Katz, 1969), it was shown that Ca^{2+} need be present only at the moment of invasion of the nerve terminal by the action potential. Changing the

timing of the influx of Ca^{2+} or the amount of influx (by changing the extracellular Ca^{2+} concentration or blocking Ca^{2+} channels with divalent cations such as Co^{2+} and Mn^{2+}) will change the timing and amplitude of the synaptic response. Similarly, an ionophore that lets Ca^{2+} ions flow across the membrane will cause a sustained increase in intracellular Ca^{2+} and the sustained outpouring of quanta of transmitter. It is also possible to fill a terminal with a caged form of the Ca^{2+} ion; that is, with Ca^{2+} ions bound up within a chemical carrier. A flash of light can rearrange this carrier, uncaging the Ca^{2+}, and thereby cause an abrupt increase in cytosolic Ca^{2+} and a sudden increase in secretion as well. Finally, loading the terminal with Ca^{2+} chelators to prevent increases in cytosolic Ca^{2+} can prevent release. Thus a rise in intracellular Ca^{2+} is the essential trigger for vesicle fusion.

Vesicles are Released by Calcium Microdomains

The opening of Ca^{2+} channels during a single action potential allows in enough Ca^{2+} to raise the concentration in the bouton from approximately 100 to 110 nM. How can a mere 10% change in the concentration cause a 1000-fold shift in the activity of the fusion machinery? Furthermore, because Ca^{2+} pumps and exchangers on the plasma membrane work relatively slowly, this small rise in Ca^{2+}, which can be detected with fluorescent Ca^{2+} indicator dyes, persists for hundreds of milliseconds after the action potential. Why then does transmitter secretion occur only for a millisecond or so? The answer to both of these questions lies in the fact that the release mechanism is *not* responding to the general concentration of Ca^{2+} in the bouton, but rather to the concentration in a *microdomain* in the immediate vicinity of the Ca^{2+} channels (Fig. 7.3). In these microdomains, very high concentrations of Ca^{2+} are reached very quickly and drop rapidly to near resting levels within microseconds of the closing of the channels. Presently, there are no Ca^{2+} dyes that are sufficiently local, sensitive, and fast that one can measure the concentration changes in these microdomains directly. Instead, our understanding of these signals comes from diffusion modeling, from modifying the buffering capacity of the cytosol, and from indirect indicators of the local Ca^{2+} concentration, such as transmitter release itself (Schneggenburger & Neher, 2000).

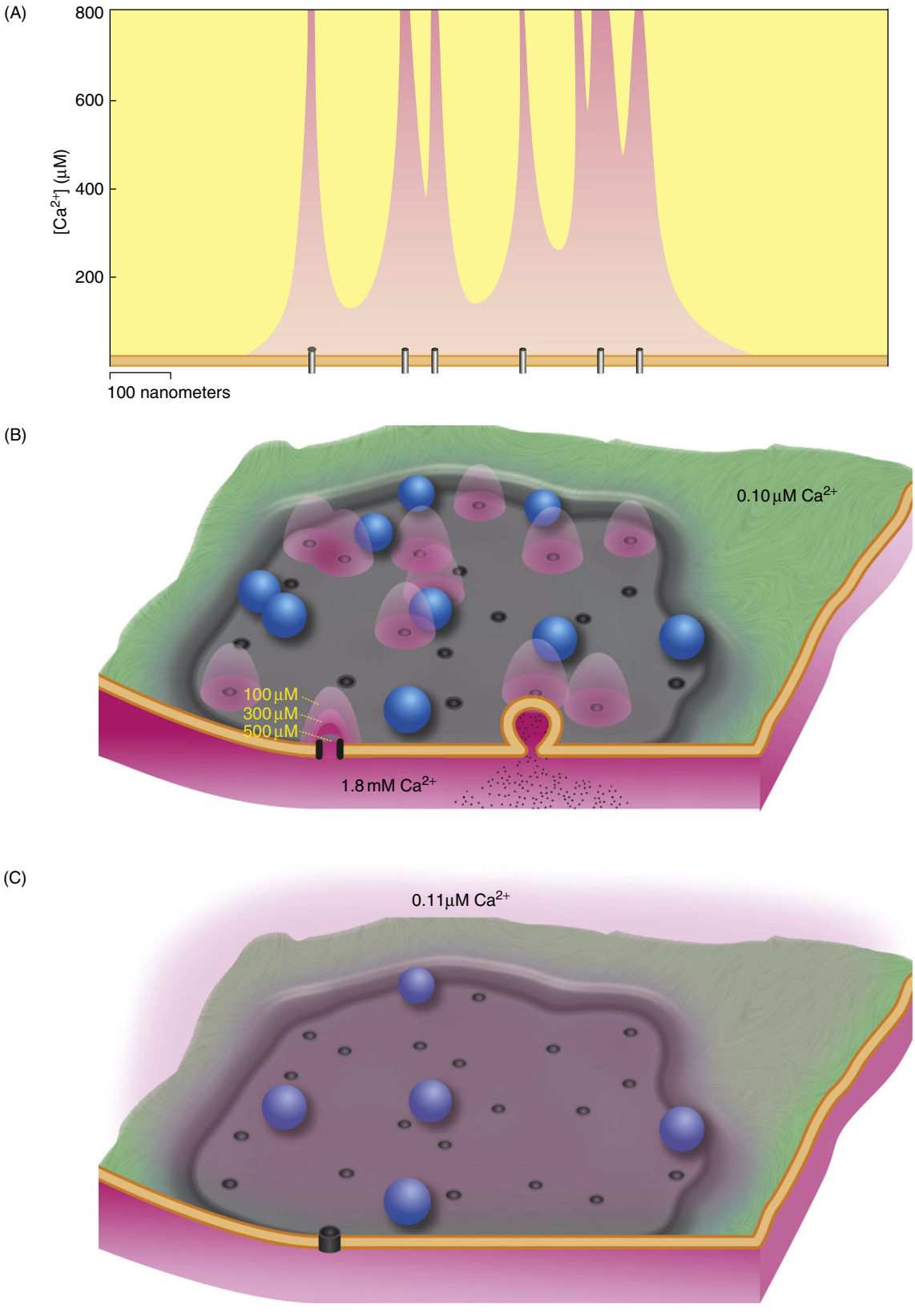
In the brief period for which the Ca^{2+} channels are open, the cytosol that lies within 100 nm of the mouth of each channel is flooded with Ca^{2+}, and the local Ca^{2+} concentration is likely to reach 100 μM or higher; the closer to the channel, the higher the concentration (Fig. 7.3A and 7.3B). The further from the channel, the more dilute the Ca^{2+} becomes and the greater the chance that it has been bound up by the high Ca^{2+}-buffering capacity of the cytosol. After the channels close, diffusion and buffering bring the Ca^{2+} concentration of the microdomains to near resting levels within a few milliseconds: the concentration gradient that existed around the mouth of the channel completely dissipates, and only the small net rise of total Ca^{2+} in the terminal remains (Fig. 7.3C). This signal, sometimes called residual Ca^{2+}, is what is detected by fluorescent indicator dyes. The effect may be compared to dumping a bucket of water into a swimming pool—a dramatic rise in water level occurs at one spot but it is very transient and gives rise to just a small net rise in the level of the entire pool.

A single active zone may have more than 100 channels in its membrane and a single vesicle may therefore be within 50 nm of as many as 10 Ca^{2+} channels. Though not all these channels will open during each action potential, more than one channel is likely to open, and therefore Ca^{2+} entering through several nearby channels can sum in overlapping microdomains to influence a vesicle (Fig. 7.3). Of the subtypes of Ca^{2+} channels that have been described, N and P/Q types of Ca^{2+} channels appear to be the most prevalent at CNS synapses, although R and L types are also reported at certain synapses.

The Exocytosis Trigger Must Have Fast, Low-Affinity, Cooperative Ca^{2+} Binding

The sensor that detects the Ca^{2+} so as to trigger release must also have special properties to achieve fast and transient exocytosis. Estimates at several synapses suggest that the Ca^{2+}-sensor has an affinity in the range of 10–100 μM, as appropriate for the relatively high concentrations of Ca^{2+} in the microdomains near channels. The on rate must be particularly fast. This is confirmed by the finding that presynaptic injection of relatively slow Ca^{2+} chelators such as ethylene glycol bis(b-aminoethyl ether)-*N*,*N*′-tetraacetic acid (EGTA) have almost no effect on transmitter release to single action potentials. Only fast Ca^{2+} chelators, such as 1,2-bis(2-aminophenoxy)ethane-*N*,*N*, *N*′,*N*′-tetraacetic acid (BAPTA), with on rates of about 5×10^8 $M^{-1}s^{-1}$, can capture Ca^{2+} ions before they bind to the secretory trigger, indicating that the on-rate of Ca^{2+} binding to this trigger is similarly fast. The off rate of Ca^{2+} dissociation from these sites must also be fast, at least 10^3 s^{-1}, to account for the rapid termination of transmitter release (0.25 ms time constant) after Ca^{2+} channels close and Ca^{2+} microdomains collapse. If microdomains did not dissipate as quickly, or if the affinity of the sensor were higher (the off rate slower), transmitter release would persist too long after the action potential had passed.

Importantly, the relationship of Ca^{2+} influx or cytosolic Ca^{2+} in microdomains to the amount of release is not linear at most synapses. By plotting this relationship on a log-log plot, a Hill coefficient can be determined, and values as high as 3 or 4 commonly are obtained. This


(A)
800
600
400
200
[Ca2+] (μM)
100 nanometers
(B)
0.10 μM Ca2+
100 μM
300 μM
500 μM
1.8 mM Ca2+
(C)
0.11μM Ca2+

value implies a high cooperativity in the action of Ca^{2+} inside the terminal, perhaps because multiple Ca^{2+}-binding sites must be occupied in order to trigger release efficiently (Schneggenburger & Neher, 2000). The interesting physiological consequence of this steep relationship is that small modulations of Ca^{2+} channels that increase or decrease Ca^{2+} influx can have a large effect on the strength of a synapse. For example, at a synapse from a parallel fiber onto a Purkinje cell in the cerebellum, when the activation of modulatory presynaptic GABA receptors inhibits Ca^{2+} channels sufficiently to reduce Ca^{2+} influx by 25%, the amplitude of transmission decreases by 70%.

Summary

Chemical synapses permit one neuron to rapidly and effectively excite or inhibit the activity of another cell. A diversity of transmitters and receptors allows varied postsynaptic responses. The packaging of transmitter into vesicles and its release in quanta enable a single action potential to secrete hundreds of thousands of molecules of transmitter almost instantaneously onto another cell. Ca^{2+} acts as an intracellular messenger tying the electrical signal of presynaptic depolarization to vesicle release. At fast synapses, Ca^{2+} enters through clusters of channels near docked synaptic vesicles in active zones. This Ca^{2+} acts at extremely short distances (tens of nanometers) in remarkably little time (200 μs) and at very high local concentrations (~100 μM), in calcium microdomains, by binding cooperatively to a low-affinity receptor with fast kinetics to trigger exocytosis. When Ca^{2+} channels close, these microdomains of high Ca^{2+} return to near resting concentrations quickly and the evoked response is terminated. Speed, efficiency, and flexibility are the hallmarks of this process.

FIGURE 7.3 Microdomains with high Ca^{2+} concentrations form in the cytosol near open Ca^{2+} channels and trigger the exocytosis of synaptic vesicles. (A) In this adaptation of a model of Ca^{2+} dynamics in the terminal, a set of Ca^{2+} channels is spaced along the x axis, as if in a cross section of a terminal. The channels have opened and, while they are open, the cytosolic Ca^{2+} concentration (y axis) is spatially inhomogeneous. Near the mouth of the channel, the influx of Ca^{2+} drives the local concentration to as high as 800 μM, but within just 50 nm of the channel, the concentration drops off to 100 μM. Channels are spaced irregularly but are often sufficiently close to one another that their clouds of Ca^{2+} can overlap and sum. (B) In the active zone (gray), an action potential has opened a fraction of the Ca^{2+} channels, and microdomains of high cytosolic Ca^{2+} (pink) arise around these open channels as Ca^{2+} flows into the cell. In the rest of the cytoplasm, the Ca^{2+} concentration is at resting levels (0.10 μM), but within these microdomains, particularly near the channel mouth, Ca^{2+} concentrations are much higher, as in (A). Synaptic vesicles docked and primed at the active zone may come under the influence of one or more of these microdomains and thereby be triggered to fuse with the membrane. (C) A few milliseconds after the action potential, the channels have closed and the microdomains have dispersed. The overall Ca^{2+} concentration in the terminal is now slightly higher (0.11 μM) than before the action potential. If no other action potentials occur, the cell will pump the extra Ca^{2+} out across the plasma membrane and restore the initial condition after several 100 ms.

MOLECULAR MECHANISMS OF THE NERVE TERMINAL

To release neurotransmitter in response to an action potential, a synaptic vesicle must fuse with the plasma membrane with great rapidity and fidelity, and thus the synapse requires an effective and well-regulated molecular machine. This machine must include the means to load the vesicle with transmitter, to dock the vesicle near the membrane so that it can fuse with a short latency, to define a release site on the plasma membrane, and to cause the release itself. Additionally, a reserve of synaptic vesicles must be held near the active zone (e.g., Fig. 7.1A) and those vesicles must be recruited to the plasma membrane as needed. The number of vesicles that are ready to fuse must be strictly determined, and the protein and lipid components of the vesicle must be recycled to form a new vesicle after fusion has occurred. For each of these processes, a molecular understanding remains incomplete, but rapid progress in this field has resulted in considerable headway.

Most Neurons Require a Cycle of Membrane Trafficking

Active neurons are in constant need of transmitter-filled vesicles ready to release their contents. A bouton in the CNS, for example, may contain a store of 200 vesicles, but if it releases even one of these with each action potential and if the cell is firing at an unexceptional rate such as 5 Hz, the store of vesicles would be consumed within less than a minute. Transport of newly synthesized vesicles from the cell body would be far too slow to support such a demand and the axonal traffic needed to supply an entire arbor of nerve terminals would be staggering.

In short, active neurons need an efficient mechanism to recycle and reload vesicles within the terminal. The exception is peptidergic neurons because peptides must be synthesized in the endoplastic reticulum (ER), chiefly in the cell body. Not surprisingly, therefore, peptide vesicles are released at very low rates. For most neurotransmitters, however, the exocytosis of a synaptic vesicle is followed rapidly by endocytosis and within approximately 30 s the vesicle is again available for release. This pathway is sometimes referred to as the exo-endocytic cycle (Fig. 7.2). If we start with a transmitter-filled vesicle in the cytosol, we can outline its progression through this cycle. The vesicle will be mobilized from the reserve pool

in the cytosol to the readily releasable pool by translocating to the active zone and becoming "docked" at the plasma membrane. Biochemical priming steps may occur at this point that will enable the vesicle to fuse within microseconds once the Ca^{2+} signal is given. Exocytosis often involves a complete merging of the vesicle membrane with the plasma membrane, though some research supports the coexistence of a second pathway involving only a transient connection of the two. In the latter case (see later), the vesicle, having emptied its contents, may return directly to the vesicle pool and be refilled with transmitter. If fusion is complete, however, the vesicle needs to be reformed with a clathrin cage and pinched off the plasma membrane. In some cases, these endocytosed vesicles seem to fuse with endosomes or large membranous sacs called cisternae, and mature synaptic vesicles then bud from this compartment. In other cases, the mature vesicle may be formed directly from the plasma membrane. In either case, the vesicle must be reloaded with transmitter prior to its reuse. The pathways for neurotransmitters are discussed in Chapter 6.

Histological tracers have provided a means to follow the vesicle cycle. One such tracer is the enzyme horseradish peroxidase (HRP), which could be placed in the extracellular solution and which would be internalized into intracellular compartments, including synaptic vesicles, when the synapse was stimulated. Another set of useful tracers are lipophilic dyes, including FM1-43, which partition into the surface membrane of the cell and are similarly internalized as membrane vesicles endocytosis. When the dye is washed from the extracellular space, the surface destains, but dye remains trapped in internal compartments. Loading of this dye upon stimulation and subsequent release of the dye with further stimulation has been an important method for examining the kinetics of the cycle.

A defect in any single step in the exo-endocytic cycle will halt transmitter release. Moreover, each step in this cycle represents a potential control point for modulating the efficacy of the synapse. For example, regulation by Ca^{2+} and other second messenger systems is known to affect the docking, fusing, and recycling of vesicles at some synapses and is also likely to regulate the balance between reserve stores and those vesicles actively engaged in the exo-endocytic cycle. Understanding the mechanisms of this modulation is an important goal and will certainly require the detailed understanding of the fundamental machinery itself.

The Fusion Mechanism is Rapid

As discussed earlier in this chapter, the delay between the arrival of an action potential at a terminal and the secretion of the transmitter can be less than 200 μs. This places severe constraints on the fusion mechanism. Vesicles must already be present at the release sites, as there is no time to mobilize them from a distance. A catalytic cascade during fusion, such as that involved in phototransduction or in excitation-contraction coupling in smooth muscle, would also be far too slow for excitation-secretion coupling at the nerve terminal. Models therefore favor the idea that a fusion-ready complex of the vesicle and plasma membrane is preassembled at release sites and that Ca^{2+} binding need only trigger a simple conformation change in this complex to open a pathway for the transmitter to exit the vesicle.

Because the volume of the synaptic vesicle is small, the diffusion of transmitter from the vesicle proceeds almost instantaneously as soon as a pore has opened up between the vesicle lumen and the extracellular space. This structure is referred to as the fusion pore, but its biochemical nature is unknown. Thus, the time-critical steps come between the influx of Ca^{2+} and the formation of the fusion pore, and these steps establish the latency between arrival of the action potential and transmitter diffusion into the cleft. Any further step, such as the complete merging of the vesicle and plasma membrane, can occur on a slower time course, after the transmitter has left the vesicle. The tethering of the vesicle at the release site (often called docking) and any biochemical events that need to occur in order for the vesicle to reach the fusion-ready state (often called priming) can also be slower. Docking and priming a vesicle cannot be *too* slow, however; a central nervous system (CNS) synapse is estimated to have 2 to 20 vesicles in this fusion-ready state; therefore, if a synapse is to respond faithfully to a sustained train of action potentials, it must be able to replace the fusion-ready vesicles with a time course of seconds. The rate at which this occurs may determine some of the dynamic properties of the synapse.

The short latency of transmission would seem to preclude the involvement of ATP hydrolysis at the fusion step because such a reaction would be too slow. Indeed, fusion can proceed without ATP present and thus, if energy is needed to fuse the membranes, it must be stored in the fusion-ready state of the vesicle-membrane complex and released upon addition of Ca^{2+}.

Transmitter Release is a General Cell Biological Question with Particularly Neuronal Refinements

Exocytosis and endocytosis are not unique to neurons; these processes go on in every eukaryotic cell. Moreover, exocytosis itself is only one representative of a general class of membrane-trafficking steps in which one membrane-bound compartment fuses with another. Other examples would include transport from the ER to

the Golgi, or transport from endosomal compartments to lysosomes. In each case, the same biophysical problem must be overcome to bring the membranes close together and drive fusion.

At present, it appears that all the membrane-trafficking steps within the cell use a similar set of proteins to accomplish this task (Fig. 7.6). Indeed, it appears that representatives of this core set of proteins are present on every trafficking vesicle or target membrane and are required for every fusion step, whether it be in yeast, an epithelial cell, or a neuron. This discovery, which grew from the conjunction of independent studies of different model systems (Bennett & Scheller, 1993), has led to the exciting hypothesis that all intracellular membrane fusions will be united by a single and universal mechanism. This has been a great boon to cell biology: experiments in one model system, for example, the highly developed genetic analysis of membrane fusion in yeast, could be absorbed into neuroscience. Similarly, the abundance of synaptic vesicles for biochemical analysis and the unparalleled precision of electrophysiological assays of single vesicle fusions could inform the understanding of other cellular events.

The disparate membrane fusion events within a cell, however, are not identical in their mechanisms. Different membrane fusions present variations on a common theme; the fundamental process is adapted to the specific requirements of each physiological step. The need for extremely tight control of fusion in the nerve terminal—for rapid rates of fusion with very short latencies to occur in brief bursts at precise points on the plasma membrane—must require some important differences in how the synaptic isoforms of these proteins function compared to the isoforms involved in general cellular traffic. The fusion of peptide-containing vesicles in nerve terminals appears to follow different rules from those that govern the release of small-molecule transmitters and must therefore entail mechanistic differences. Moreover, recent findings suggest that the spontaneous release of vesicles at nerve terminals may have important mechanistic differences from the release that is triggered by an action potential. Thus, for a full understanding of the mechanisms of the nerve terminal, we will need to understand not only the common fundamentals of vesicle fusion, but also the regulatory mechanisms that may be specialized for different classes of transmission and determine the individual characteristics of a synapse.

Synaptic Proteins Have Been Identified from Purified Vesicles

Synaptic vesicles are abundant in nervous tissue and, due to their unique physical properties (uniform small diameter and low buoyant density), they can be purified to homogeneity by subcellular fractionation techniques. As a result, at a biochemical level, synaptic vesicles are among the most thoroughly characterized organelles. Indeed, with the advent of highly sensitive mass spectrometric methods of protein identification, a very comprehensive and quantitative analysis of the proteins in purified synaptic vesicles has been possible (Takamori et al., 2006). The protein compositions of synaptic vesicles from different sources are remarkably similar, demonstrating that many of the proteins present on the synaptic vesicle membrane perform general functions that are not restricted to a single class of transmitter.

Our knowledge of many of the important proteins in vesicle fusion commenced with the characterization of purified synaptic vesicles from which individual proteins could be isolated and subsequently cloned. The major constituents of the synaptic vesicle are shown in Figure 7.4. For some of these proteins, there are well-established functions, but for others, the functions remain uncertain (Table 7.1). The proton transporter, for example, is an ATPase that acidifies the lumen of the vesicle. The resulting proton gradient provides the energy by which transmitter is moved into the vesicle. This task is carried out by another vesicular protein, a vesicular transporter that allows protons to move down their electrochemical gradient and out of the vesicle in exchange for transmitter that moves into the lumen. Several vesicular transporters are known, all structurally related, and by their substrate selectivity they help equip a terminal for release of the appropriate transmitter (see later). Two other large proteins with multiple transmembrane domains are abundant in vesicles: synaptic vesicle protein 2 (SV2) and synaptophysin. The mechanism of action of these proteins remains elusive despite extensive biochemical and genetic

TABLE 7.1 Function of Synaptic Vesicle Proteins

Protein	Function
Proton pump	Generation of electrochemical gradient of protons
Vesicular transmitter transporter	Transmitter uptake into vesicle
VAMP/synaptobrevin	Component of SNARE complex; acts in a late, essential step in vesicle fusion
Synaptotagmin	Ca^{2+}-binding trigger for fusion and component of vesicle docking at release sites via interactions with SNARE complex and lipid; promotes clathrin-mediated endocytosis by binding AP-2 complex
Rab3	Possible role in regulating vesicle targeting and availability
Synapsin	Likely to tether vesicle to actin cytoskeleton
Cysteine string protein	Promotes reliable coupling of action potential to exocytosis
SV2	Unknown
Synaptophysin	Unknown, endocytosis?

characterization. Other vesicular proteins are discussed later in this chapter.

Transmitter release depends on more than just vesicular proteins, however; proteins of the plasma membrane and cytoplasm are also important. In many cases, the identification of these additional components or an appreciation of their importance to the synapse derived from investigations of vesicular proteins. Two examples will serve to illustrate this point. The first example begins with synaptotagmin, an integral membrane protein of the synaptic vesicle that was purified from synaptic vesicles and cloned. The portion of synaptotagmin that extends into the cytoplasm (the majority of the protein; see Fig. 7.4) was subsequently used for affinity column chromatography. In this manner, a protein called syntaxin was identified as a synaptotagmin-binding protein. This protein resides in the plasma membrane and is now appreciated as one of the critical players in vesicle fusion (see later). The interactions of syntaxin with many other synaptic proteins have been studied intensively to shed light on fusion and its regulation. Thus, subsequent to the isolation of an abundant vesicular protein, biochemical assays have led to a fuller picture.

A second and similar example (Gonzalez & Scheller, 1999) began with the realization that an abundant small GTP-binding protein, rab3, was present on the vesicle surface. This protein, discussed further later, may regulate vesicle availability or docking to release sites. Subsequently, rabphilin was identified on the basis of its affinity for the GTP-bound form of rab3. Rabphilin lacks

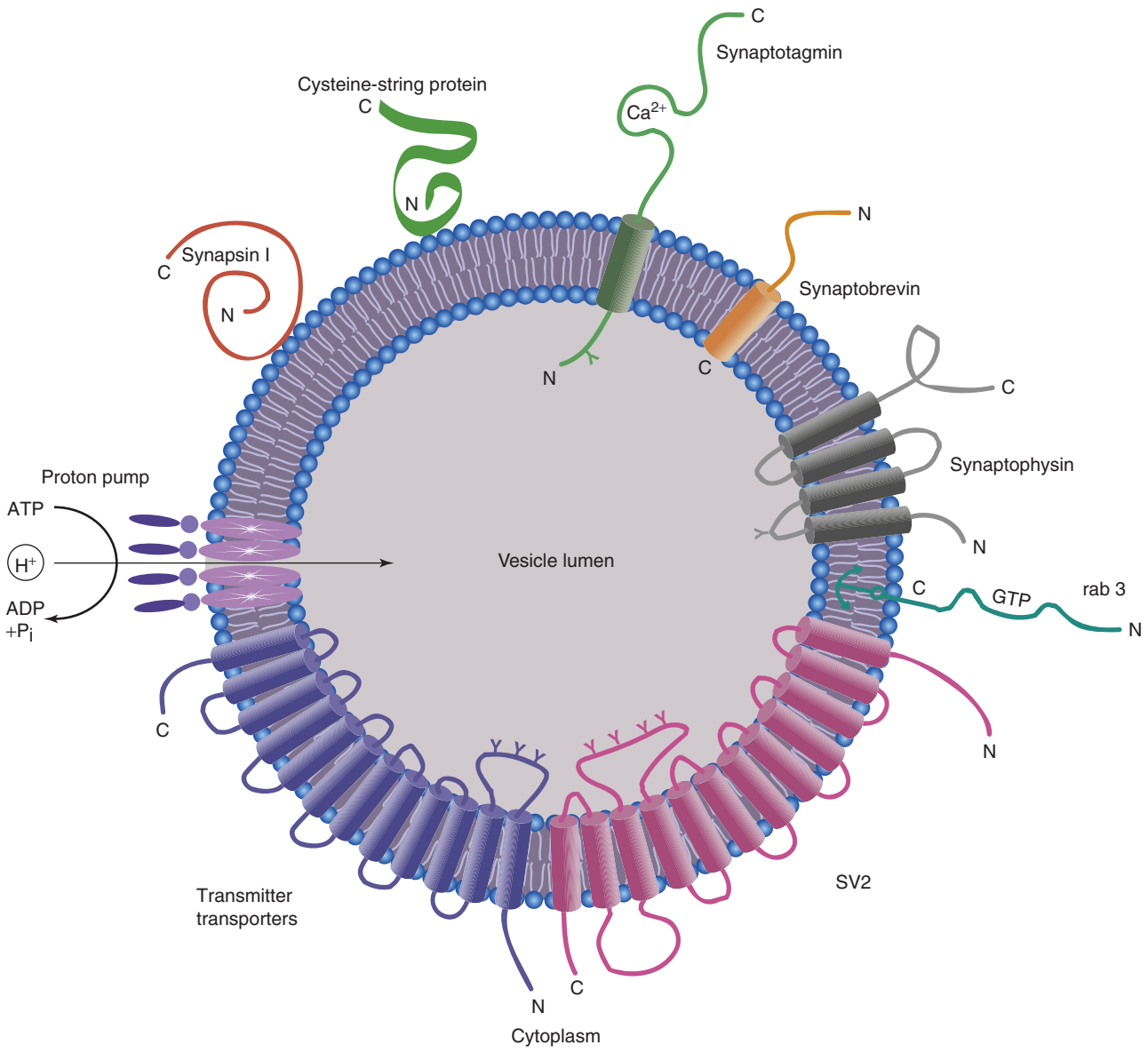


FIGURE 7.4 Schematic representation of the structure and topology of major synaptic vesicle membrane proteins (see also Table 8.1).

a transmembrane domain but is recruited to the surface of the synaptic vesicle by binding to rab3. Rabphilin may have a role in modulating transmission, particularly in mossy fiber terminals of the hippocampus. Still later, many additional rab3-binding proteins were identified, including RIM, a component of the active zone. Rab3 can exist in either GTP- or GDP-bound states, and additional factors that regulate these states were also identified: a GDP dissociation inhibitor (GDI), a GDP/GTP exchange protein (GEP), and a GTPase activating protein (GAP). Thus, from the identification of a synaptic vesicle protein, an array of additional factors has come to light, all of which are likely to figure in the exo-endocytic cycle.

Genetic Screens Have Led to the Identification of Synaptic Proteins

Genetic screens have provided an independent method for identifying the machinery of transmitter release. One of the most fertile screens was carried out not in the nervous system, but in yeast. Because membrane trafficking in yeast is closely parallel to vesicle fusion at the terminal, mutations that alter the secretion of enzymes from yeast can be a springboard for the identification of synaptic proteins. In the early 1980s, a series of such screens was carried out, and a collection of over 50 mutants was obtained. Many of these caused post-Golgi vesicles to accumulate in the cytoplasm and thus appeared to block a late stage of transport, such as the targeting or fusion of these vesicles at the plasma membrane. Screens for suppressors and enhancers of these secretion mutations uncovered further components. Subsequently, excellent *in vitro* assays have been established in which to study the fusion of vesicles derived from yeast with their target organelles. Among the secretion mutants and their interacting genes were homologues of some of the proteins discussed earlier: sec4 encodes a small GTP-binding protein like rab3, and Sso1 and Sso2 encode plasma membrane proteins that are homologues of syntaxin. The sec1 gene encodes a soluble protein with a very high affinity for Sso1, and the mammalian homologue of this protein, n-sec1, is bound tightly to syntaxin in nerve terminals and has an essential function in transmission. The convergence of neuronal biochemistry with yeast genetics has built a strong case for a common mechanism of vesicle fusion.

Yeast is not the only organism in which a genetic screen uncovered an important protein for the synapse: the unc-13 mutation of *Caenorhabditis elegans*, for example, identified a component of the active zone membrane that is important in priming vesicles for fusion and in modulation of the synapse.

Genetics has further contributed to our understanding of synaptic proteins by allowing tests of the significance of an identified protein for synaptic transmission. Such studies have been carried out in *C. elegans*, *Drosophila*, and mice, and can reveal either an absolute requirement for the protein (as in the case of syntaxin mutants in *Drosophila*) or relatively subtle effects (as in the case of rab3 mutations in mice).

From biochemical purifications, *in vitro* assays, genetic screens, and fortuitous discoveries, a list of nerve terminal proteins has been assembled (Tables 7.1 and 7.2). The manner in which these proteins coordinate the release of transmitter, as well as all the other cell biological functions of the exo-endocytic cycle, remains uncertain, but a consensus has emerged in recent years that puts one set of proteins at the core of the vesicle fusion.

TABLE 7.2 Additional Proteins Implicated in Transmitter Release

Protein	Function
Syntaxin	SNARE protein present on plasma membrane (and on synaptic vesicles to a lesser extent); forms core complex with SNAP-25 and VAMP/synaptobrevin; essential for late step in fusion
SNAP-25	SNARE protein present on plasma membrane (and on synaptic vesicles to a lesser extent); forms core complex with syntaxin and VAMP/synaptobrevin; essential for late step in fusion
Nsec-l/munc-18	Syntaxin-binding protein required for all membrane traffic to the cell surface
Complexin	Syntaxin-binding protein; may stabilize an intermediate in core complex formation
Snapin	Binds SNAP-25; associated with synaptic vesicles; unknown function
NSF	ATPase that disassembles SNARE complex; likely to disrupt complexes after exocytosis
a-SNAP	Cofactor for NSF in SNARE complex disassembly
unc-13/munc-13	Active zone protein; vesicle priming for release; modulation of transmission by diacyl glycerol and Protein Kinase C
Rabphilin	C2 domain protein; Ca^{2+}-binding protein; binds rab3 and associates with synaptic vesicle; modulation of transmission (?)
DOC2	Ca^{2+}-binding C2 domain protein; binds munc-18 and SNAREs; regulates spontaneous fusions and asynchronous release
RIM1 and related proteins	Active zone proteins; bind rab3; modulation of transmission
Piccolo	Likely scaffolding protein at active zones
Bassoon	Likely scaffolding protein at active zone
Exocyst (sec6/8 complex)	Marks plasma membrane sites of vesicle fusion; not needed for synaptic vesicle fusion

SNAREs and the Core Complex are Key to Membrane Fusions

Three synaptic proteins, vesicle-associated membrane protein (VAMP), syntaxin, and the synaptosome-associated protein of 25 kDa (SNAP-25), are capable of forming an exceptionally tight complex with one

another that generally is referred to as either the *core complex* or the *SNAP receptor (SNARE) complex* (Sollner, Bennett, Whiteheart, Scheller, & Rothman, 1993a, 1993b; Sutton, Fasshauer, Jahn, & Brunger, 1998). The interaction of these three proteins is essential for synaptic transmission and is likely to lie very close to or indeed at the final fusion step of exocytosis (Figs. 7.5 and 7.6). What are these proteins?

VAMP (also called synaptobrevin) was among the first synaptic vesicle proteins to be cloned. It is anchored to the synaptic vesicle by a single transmembrane domain and has a cytoplasmic domain that contributes a coiled-coil strand to the core complex. Syntaxin has a very similar structure but is located primarily in the plasma membrane (although some is present on vesicles as well). SNAP-25 is also a protein primarily of the plasma membrane but, unlike the others, lacks a transmembrane domain and instead is anchored in its central region by acylations. SNAP-25 contributes two strands to the SNARE coiled coil. The interactions of these proteins can be envisioned as closely juxtaposing the two membranes meant to fuse. The proteins of this complex are archetypes of a class of membrane-trafficking protein collectively called *SNAREs*. Vesicle-associated proteins, such as VAMP, are referred to as *v-SNAREs*, and those of the target membrane, such as SNAP-25 and syntaxin, are referred to as *t-SNAREs*.

A SNARE complex is found at each membrane-trafficking step within a eukaryotic cell (Fig. 7.6). Through a combination of SNARE proteins on the opposing membranes, a four-stranded coiled-coil is formed. Alongside the example of the synapse in Figure 7.6A, three analogous cases are shown from exocytosis in yeast (Fig. 7.6B), the fusion of late endosomes with one another (Fig. 7.6C), and the fusion of vesicles that form the yeast vacuole (Fig. 7.6D).

Abundant genetic and biochemical data argue for an essential role of SNAREs, and the combination of yeast genetics, *in vitro* assays, biochemical analyses, and synaptic physiology has had a synergistic effect in advancing the field. For example, the discovery that secretory mutations in yeast were homologues of synaptic SNAREs provided some of the first functional data on these proteins (Bennett & Scheller, 1993), whereas data from synapses first placed the proteins on vesicles and the plasma membrane. In addition, the crystal structure was solved first for synaptic proteins (Sutton et al., 1998). Because transmembrane domains and, of course, the membranes themselves were absent from the crystallized samples, it was not possible to conclude from the structure alone that SNAREs form a bridge between the membranes rather than forming complexes within the same membrane. Instead, an *in vitro* assay with purified yeast vesicles established this point.

Some of the strongest evidence for an essential role of SNAREs at the synapse has come from the study of a potent set of eight neurotoxins produced by clostridial bacteria. These toxins (tetanus toxin and the family of related botulinum toxins) have long been known to block the release of neurotransmitter from the terminal. The discovery that they do so by proteolytically cleaving individual members of the SNARE complex provided neurobiologists with a set of tools with which to probe SNARE function (Fig. 7.5).

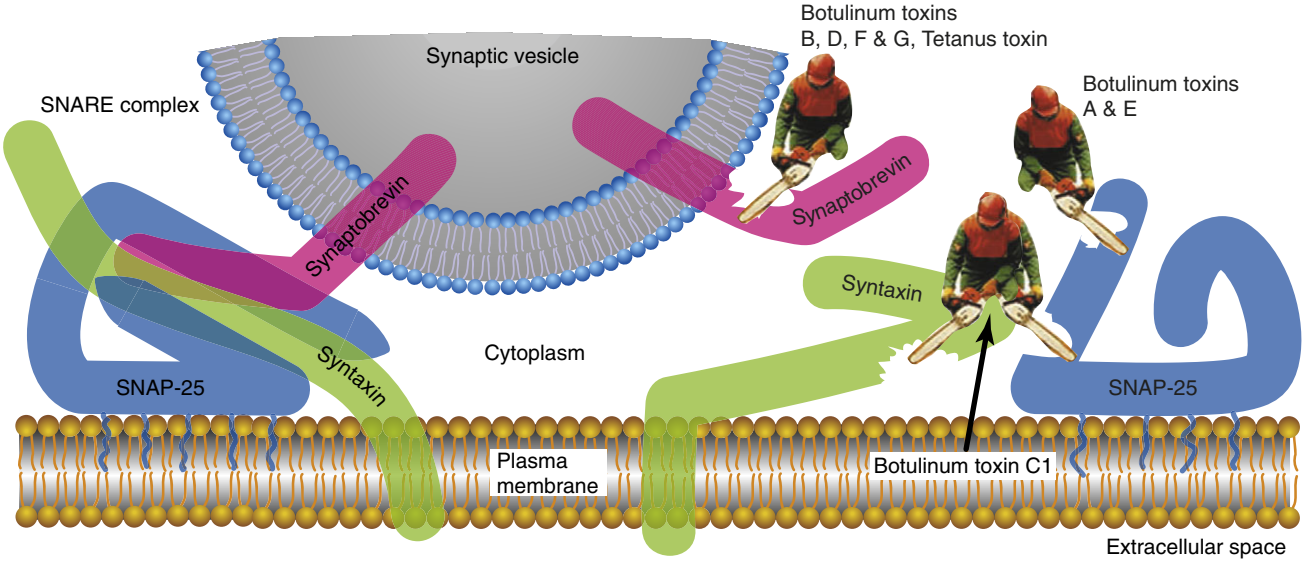


FIGURE 7.5 SNARE proteins and the action of clostridial neurotoxins. The SNARE complex shown at the left brings the vesicle and plasma membranes into close proximity and likely represents one of the last steps in vesicle fusion. Vesicular VAMP, also called synaptobrevin, binds with syntaxin and SNAP-25 that are anchored to the plasma membrane. Tetanus toxin and the botulinum toxins, proteases that cleave specific SNARE proteins as shown, can block transmitter release.

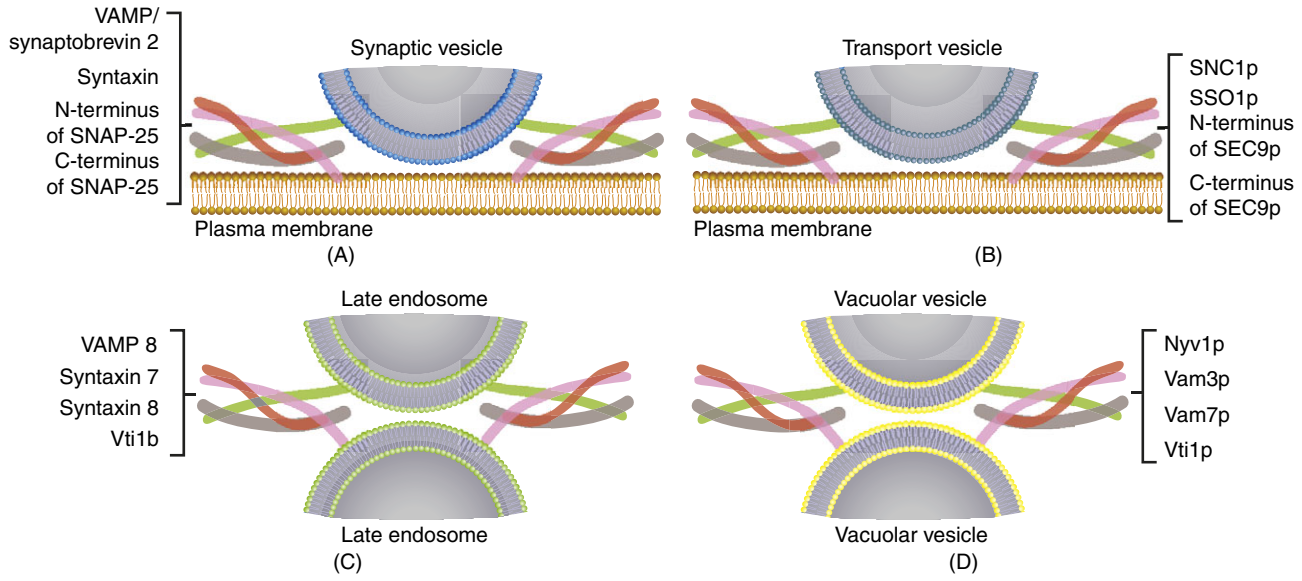


FIGURE 7.6 Neurotransmitter release shares a core mechanism with many membrane fusion events within eukaryotic cells. The fusion of synaptic vesicles (A) is driven by a particular complex of four coiled-coil domains contributed by three different proteins. Exocytosis in yeast (B), the fusion of late endosomes in mammalian cells (C), and the fusion of vacuolar vesicles in yeast (D) exemplify the closely related four-stranded coiled-coil complexes required to drive fusion in other membrane-trafficking steps.

Each of the toxins comprises a heavy and a light chain that are linked by disulfide bonds. The heavy chain binds the toxin to surface receptors on neurons, thereby enabling the toxin to be endocytosed. Once inside the cell, the disulfide bond is reduced and the free light chain enters the cytoplasm of the cell. This light chain is the active portion of the toxin and is a member of the Zn^{2+}-dependent family of proteases. The catalytic nature of the toxin accounts for its astonishing potency; a few tetanus toxin light chains at a synapse can suffice to proteolyse all the VAMP/synaptobrevin, thereby shutting down transmitter release. The toxins are highly specific, recognizing unique sequences within an individual SNARE protein, as summarized in Fig. 7.5. VAMP/synaptobrevin can be cleaved not only by tetanus toxin but by the botulinum toxins of types B, D, F, and G, and each toxin cleaves at a different peptide bond within the structure. SNAP-25 is cleaved by botulinum toxins A and E, but again at different sites from one another. Botulinum toxin C1 cleaves both syntaxin and, less efficiently, SNAP-25.

What precisely is the function of SNARE proteins in promoting transmitter release? How does the assembly of this complex relate to membrane fusion? Studies with botulinum toxins, yeast mutants, mutants of *Drosophila* and *C. elegans*, and permeabilized mammalian cells all place SNAREs late in the process. Synapses that lack an individual SNARE—for example, synapses whose VAMP has been mutated or cleaved by toxin—have the expected population of synaptic vesicles, and these vesicles accumulate at active zones in the expected manner. Indeed, as judged by electron microscopy, there may even be an excess of vesicles in close proximity to the plasma membrane at the active zone. Yet these synapses are incapable of secreting transmitter in response to an action potential. Thus, SNAREs appear to be essential for a step that comes after the arrival of the vesicle to the release site, but before the fusion pore opens and transmitter can diffuse into the cleft. The details of how the complex functions, however, are less certain.

One attractive model is that the energy released by the formation of this very high-affinity complex is used to drive together the two membranes. A loose complex of the SNARES would form, probably involving the N-terminal portions of the SNARE proteins and this might represent a fusion-ready but stable intermediate state. The SNARE proteins would then "zipper up" into the full core complex structure in response to the triggering signal Ca^{2+} signal. The zippering up will pull the membranes together and this in turn will allow a fusion pore to form that will connect the vesicle lumen to the extracellular space. Vesicles consisting of only lipid bilayers and SNAREs can fuse *in vitro*, suggesting that this machinery may be sufficient for the fusion step although *in vivo* additional proteins are required.

One additional function may reside with SNARE proteins: identification of an appropriate target membrane. Within a cell, there are myriad membrane compartments with which a transport vesicle can fuse: how then is specificity achieved? The great diversity of SNAREs (Fig. 7.6) may account for some of this specificity because not all combinations of vesicle-associated and

target-associated SNAREs (v- and t-SNAREs) will form functional complexes. This potential mechanism, however, is likely to be only a part of the story. Particularly in the nerve terminal, it appears that synaptic vesicles can find the active zone even in the absence of the relevant SNAREs—for example, when a SNARE has been cleaved by a clostridial toxin or in a mutant lacking a SNARE. In addition, the t-SNAREs' syntaxin and SNAP-25 can be present along the entire axon and thus are inadequate to explain the selective release of transmitter at synapses and active zones.

NSF—An ATPase for Membrane Trafficking

At some point as vesicles move through their exo-endocytic cycle, energy must be added to the system. *N*-Ethylmaleimide sensitive factor (NSF), an ATPase involved in membrane trafficking, is one likely source. NSF was first identified as a required cytosolic factor in an *in vitro* trafficking assay. The importance of NSF was confirmed when it was found to correspond to the yeast sec18 gene, an essential gene for secretion. NSF hexamers bind a cofactor called a-SNAP (Soluble NSF Attachment Protein), or sec17, and this complex in turn can bind to the SNARE complex. When Mg-ATP is hydrolyzed, the SNARE complex is disrupted into its component proteins (Sollner et al., 1993a, 1993b). If SNARE complexes form between VAMP, syntaxin, and SNAP-25 all in the same membrane, these futile complexes can be split apart by NSF so that productive complexes bridging the membrane compartments can be formed. After fusion, the tight SNARE complex needs to be disrupted so that the VAMP can be recycled to synaptic vesicles while the other SNAREs remain on the plasma membrane. If, indeed, the energy of forming a tight SNARE complex is part of the energy that drives fusion, NSF, by restoring the SNAREs to their dissociated, high-energy state, will be an important part of the energetics of membrane fusion.

Docking and Priming Vesicles Prepares Them for Fusion

As discussed earlier and outlined in Figure 7.2, many preparatory and regulatory steps may precede the action of SNAREs and the membrane fusion step at which SNAREs appear to act. These preparatory steps must tether the vesicle at an appropriate release site in the active zone and hold the vesicle in a fusion-ready state. These mechanisms remain among the most obscure of the processes involved in exocytosis, but there is a growing list of proteins that may participate (Table 7.2). One example is the syntaxin-binding protein munc-18 (also called nsec1), the neuronal homologue of the product of the yeast secl gene. Mutations of this protein prevent trafficking in both yeast and higher organisms. It appears likely that munc18 serves two functions and binds to SNARE complexes in more than one fashion: it may promote membrane fusion by binding to syntaxin in a manner that will subsequently promote SNARE complex formation, but it may also be a negative regulator, keeping syntaxin inert until an appropriate vesicle or signal displaces it and allows a SNARE complex to form. The sequencing of when and how it binds and unbinds remains a topic of study, but munc18 is clearly essential for transmitter release.

Rab3, a small GTP-binding protein mentioned earlier, is another protein for which a priming or regulatory role often is invoked at the synapse. In yeast, and at other membrane-trafficking steps within mammalian cells, rab proteins have essential roles. They appear to help a vesicle to recognize its appropriate target and begin the process of SNARE complex formation. At the synapse, however, the significance of rab3 is still uncertain. Although it is clearly associated with synaptic vesicles, genetic disruption of rab3 has a surprisingly slight phenotype and causes only subtle alterations in synaptic properties. Whether this is due to additional, redundant rab proteins or whether the rab family has been relegated to a more minor role at the synapse remains to be determined. Interestingly, in *Drosophila*, the chief phenotype of *rab3* mutants is in the assembly of the proteins of the active zone rather than in the fusion steps themselves.

Because synaptic vesicles dock and fuse specifically at the active zone, this region of the nerve terminal membrane must have unique properties that promote docking and priming. The special nature of this domain is easily discernible in electron micrographs: an area of electron-dense material can be observed opposite the postsynaptic density (Fig. 7.1). In some synapses, such as photoreceptors, hair cells, and many insect synapses, the structures are more elaborate and include ribbons, dense bodies, and T bars that extend into the cytoplasm and appear to have a special relationship with the nearby pool of vesicles. Advances in electron microscopy have allowed a more detailed look at the association of vesicles and plasma membrane at the active zone (Harlow, Ress, Stoschek, Marshall, & McMahan, 2001). At the neuromuscular junction of the frog (Fig. 7.1), the electron-dense material adjacent to the presynaptic plasma membrane is actually a highly ordered structure—a lattice of proteins that connect the vesicles to a cytoskeleton, to one another, and to the plasma membrane. The molecules that correspond to these structures are not yet known. A few proteins, however, are known to be concentrated in the active zone or in the cloud of vesicles near the active zone. These proteins, piccolo, bassoon, RIM1, DOC2, and unc-13, may be a part of the machinery that defines the active zone as the appropriate target for

synaptic vesicle fusion. In addition, these proteins, by docking and priming vesicles for release can establish the number of vesicles available for rapid fusion upon arrival of an action potential. Because the size of that pool helps determine how many vesicles actually will fuse, the proteins of the active zone are important regulatory components of the synapse.

In addition to the specializations of the active zone, additional machinery must be present to preserve a dense cluster of synaptic vesicles extending approximately 200 nm back from the active zone (Fig. 7.1). The vesicles in this domain are not likely to be releasable within microseconds of the arrival of an action potential but are instead likely to represent a reserve pool from which vesicles can be mobilized to release sites on the plasma membrane. The equilibrium between this pool and vesicles actually at the membrane may be an important determinant of the number of vesicles released per impulse, but remains poorly understood.

One protein family likely to play a role in the maintenance of the reserve pool is the synapsins, a set of peripheral membrane proteins on synaptic vesicles. Synapsins can also bind actin filaments and thus may provide a linker that tethers the vesicles in the cluster to the synaptic cytoskeleton. Disruption of this link can cause the vesicle cluster to be diminished and reduce the number of vesicles in the releasable pool of the terminal. Synapsin has attracted considerable interest because it is the substrate for phosphorylation by both cAMP and Ca^{2+}-dependent protein kinases. These phosphorylations may influence the availability of reserve vesicles for recruitment to release sites.

Ca^{2+}-Binding Proteins are Candidates for Coupling the Action Potential to Exocytosis

The most striking difference between synaptic transmission and traffic between other cellular compartments is the rapid triggering of fusion by action potentials. As discussed previously, the opening of Ca^{2+} channels and the focal rise of intracellular Ca^{2+} activate the fusion machinery. How does the terminal sense the rise in Ca^{2+}? What is the Ca^{2+} trigger and how does it open the fusion pore? Is it a single Ca^{2+}-binding protein or do several components respond to the altered Ca^{2+} concentration? Does Ca^{2+} remove a brake that normally prevents a docked, primed vesicle from fusing or does Ca^{2+} induce a conformational change that is actively required to promote fusion? How does the steep, exponential relationship of release to intracellular Ca^{2+} arise? These questions are an active area of investigation and debate.

The leading candidate for being the synaptic Ca^{2+} sensor is synaptotagmin, an integral membrane protein of the synaptic vesicle (Fig. 7.4). Synaptotagmin has a large cytoplasmic portion that comprises two Ca^{2+}-binding C2 domains, called C2A and C2B. These domains can also interact with the SNARE complex proteins and with phospholipids in a Ca^{2+}-dependent manner. It has been hypothesized that one or more of these interactions is the molecular correlate of the triggering event for fusion. Consistent with this hypothesis, mutations that remove synaptotagmin profoundly reduce synaptic transmission in flies, worms, and mice, while having little effect on or enhancing the rate of spontaneous release of transmitter. Synaptotagmin is likely to be involved in endocytosis and potentially in vesicle docking as well, which has complicated the analysis. These processes, as mentioned earlier, are also likely to be regulated by Ca^{2+}, and the Ca^{2+}-binding sites on synaptotagmin may be relevant for this regulation as well.

Recently, an intriguing model has emerged in which the fusion of membranes is arrested at an intermediate state by the binding of a small protein called complexin to the assembled SNARE complex. Complexin appears to stabilize an intermediate step in the fusion pathway, perhaps partially zippered SNARE complexes, and the unbinding of complexin through the action of synaptotagmin and Ca^{2+} is thought to allow the fusion reaction to proceed. In this model, complexin serves as a brake and the primary function of synaptotagmin in exocytosis is to release that brake. In other models, the insertion of synaptotagmin domains into the lipid bilayer is also essential for membrane destabilization and fusion pore opening (Rizo & Rosenmund, 2008). Another complication for analysis of the Ca^{2+}-sensing trigger is that there may well be more than one. Loss of synaptotagmin isoforms always leaves some residual transmitter release, although it tends not to be as synchronous or as efficiently evoked by individual action potentials. Recent experiments have indicated that another protein containing two C2 domains, DOC2, may be the sensor for this slower phase of release and may confer Ca^{2+} sensitivity on the mini frequency (Yao, Gaffaney, Kwon, & Chapman, 2011).

Transmitter Must Be Packaged into the Vesicle

To achieve the synchronous release of thousands of molecules of transmitter from the presynaptic nerve terminal, synaptic vesicles accumulate and store high concentrations of transmitter. In cholinergic neurons, for example, the concentration of acetylcholine within the synaptic vesicle can reach 0.6 M, more than 1000-fold greater than that in the cytoplasm. Two synaptic vesicle proteins mediate the uptake of transmitter: the vacuolar proton pump and a family of transmitter transporters. The vacuolar proton pump is a multisubunit ATPase that catalyzes the translocation of protons from

the cytoplasm into the lumen of a variety of intracellular organelles, including synaptic vesicles. The resulting transmembrane electrochemical proton gradient is utilized as the energy source for the active uptake of transmitter by transmitter transporters.

Transmitter uptake has been characterized in isolated synaptic vesicle preparations in which at least four types of distinct transporters have been identified: one for acetylcholine, another for catecholamines and serotonin, a third for the excitatory amino acid glutamate, and the fourth for the inhibitory amino acids GABA and glycine. At least one gene for each of these classes of transmitter transporter has now been cloned. As expected, these distinct transporters are expressed differentially by neurons. The type of transporter in a cell dictates the type of transmitter stored in the synaptic vesicles of a particular neuron, and when investigators drive the expression of a glutamate transporter, for example, in a GABA-releasing neuron, they can trick the cell into releasing glutamate.

Vesicular transporters are integral membrane proteins with 12 membrane-spanning domains that display sequence similarity with bacterial drug-resistance transporters. Synaptic vesicle transporters are clearly distinct from the plasma membrane transmitter transporters that remove transmitter from the synaptic cleft and thereby contribute to the termination of synaptic signaling. The distinguishing characteristics include their transport topology, energy source, pharmacology, and structure. The sequestration of transmitter into vesicles, in addition to its obviously essential role for transmitter exocytosis, has some further benefits to the cell. Sequestration can prevent high cytosolic concentrations of the transmitter from inhibiting the biosynthetic enzymes for transmitter synthesis and can protect the transmitter from catabolic enzymes. In addition, in some cell types, sequestration may protect the cell from damage caused by the oxidation of labile transmitters, particularly dopamine.

In contrast to small chemical transmitters, proteinaceous signaling molecules, including neuropeptides and hormones, typically are stored in granules that are larger and have a higher electron density than synaptic vesicles. The contents of these granules are not recycled at the release sites; as a result, their replenishment requires new protein synthesis followed by packaging into secretory vesicles in the cell body. Because of the slow kinetics of their release, the slow responsiveness of their postsynaptic receptors, and their inability to be recycled locally, proteinaceous signaling molecules typically mediate regulatory functions.

Endocytosis Recovers Synaptic Vesicle Components

After exocytosis, the components of the synaptic vesicle membrane must be recovered from the presynaptic plasma membrane, as discussed previously. Vesicle recycling theoretically could be accomplished by either of two mechanisms (Fig. 7.2). The first is simply a reversal of the fusion process. In this case, a fusion pore opens to allow transmitter release and then closes rapidly to reform a vesicle. Often nicknamed "kiss and run," it has the theoretical advantage that it would allow all the vesicular components to remain together on a single vesicle that would be available immediately for reloading with transmitter. The mechanism is potentially quick and energetically efficient. This mechanism is employed in some nonneuronal cells, but its relevance for the synapse is an unresolved and highly controversial issue.

The predominant pathway for synaptic vesicle recycling is more likely to be a second mechanism: endocytosis. When a vesicle fuses completely with the plasma membrane, its protein constituents become a part of the plasma membrane and at least many of them appear to diffuse laterally in the membrane, mixing with other plasma membrane components. The reformation of a vesicle requires that the appropriate proteins of the synaptic vesicle be reassembled in the proper ratios. For some proteins, including SNAREs, there are significant populations of these proteins in the surface membrane and the particular copy of the protein that is endocytosed is unlikely to be that which was just exocytosed. Other vesicular proteins may remain together in a cluster to facilitate reassembly into a vesicle. This reassembly and subsequent endocytosis, in many cases, will occur at a site near the active zone, but not at the active zone itself. The assembly of the constituents is facilitated by adaptor proteins in the cytoplasm such as AP-2 and stonin. In addition, very local changes in the composition of the lipid membrane, in particular an increase in $PI(4,5)P_{2,}$ is important for assembly at the endocytic site. In the next stage of endocytosis, the membrane begins to bud inward and a lattice of clathrin forms on it. Clathrin-coated vesicles are the hallmark of the classical endocytic pathway in every cell. Other proteins such as epsin contribute to the recruitment of clathrin and to controlling the shape of the forming vesicle. In a nerve terminal, the shape is a crucial determinant of the volume of the vesicle and therefore the amount of transmitter held. Another class of proteins, characterized by the shared structural feature of a BAR domain, are also needed to shape vesicles and in particular to convert the invaginating bud into a spherical structure with only a narrow neck connecting it to the plasma membrane. The BAR proteins include amphiphysin and endophilin and they associate with a third protein called synaptojanin which increases the efficiency of the endocytic process. The final pinching off of a clathrin-coated vesicle requires the protein dynamin, which can form a ring-like collar around the neck of an endocytosing vesicle.

Dynamin is a GTPase and the hydrolysis of GTP causes the dynamin ring to constrict and thereby drive the scission of the vesicle from the plasma membrane. The importance of dynamin is demonstrated most clearly in *Drosophila* via a temperature-sensitive dynamin mutant known as *shibire*: at the nonpermissive temperature, *shibire* flies become paralyzed rapidly due to a nearly complete depletion of synaptic vesicles from their nerve terminals. Once the components of the synaptic vesicle membrane have been recovered in clathrin-coated vesicles, recycling is completed by vesicle uncoating and, perhaps, passage through an endosomal compartment in the nerve terminal (Fig. 7.2).

Summary

The life of the synaptic vesicle involves much more than just the Ca^{2+}-dependent fusion of a vesicle with the plasma membrane. It is a cyclical progression that must include endocytosis, transmitter loading, docking, and priming steps as well. In many regards, this cell biological process shares mechanistic similarities with membrane trafficking in other parts of the cell and with simpler organisms such as yeast. Interaction of vesicular and plasma membrane proteins of the SNARE complex—VAMP/synaptobrevin, syntaxin, and SNAP-25—is an essential late step in fusion. Many other proteins have been identified that are likely to precede the action of the SNAREs, regulate the SNAREs, and recycle them. Components of the synaptic vesicle membrane, the presynaptic plasma membrane, and the cytoplasm all contribute to the regulation of synaptic vesicle function. Together, these proteins build on the fundamental core apparatus to create an astonishingly accurate, fast, and reliable means of delivering transmitter to the synaptic cleft.

QUANTAL ANALYSIS: PROBING SYNAPTIC PHYSIOLOGY

A quantitative description of the signal passing across a synapse can be a source of insight into the biophysics of transmission and its regulation. As discussed earlier, the response to an action potential in the presynaptic cell reflects the release of discrete packets of transmitter, called quanta, that are a direct consequence of the vesicular release of transmitter. A given synapse, however, will not secrete exactly the same number of vesicles in response to each stimulus. Instead, the probabilistic nature of the presynaptic apparatus typically causes the amount released to fluctuate from trial to trial in a quantal manner; that is, the postsynaptic response adopts preferred levels, which arise from the summation of various numbers of vesicle fusions (Katz, 1969). Thus, although a particular synapse may release an average of two vesicles per action potential, it may release one vesicle in response to one stimulation and three vesicles in response to the next. To the third stimulus, the synapse may fail to release any transmitter at all. From a full description of the properties of the synapse, the size of the individual quantum response can be estimated, as well as the average number of quanta released for a given presynaptic action potential. This approach, called quantal analysis, is also a source of insight into the probabilistic processes underlying transmitter release from the presynaptic terminal and into the mechanisms by which transmission can be modified by physiological, pharmacological, and pathological phenomena.

A Standard Model of Quantal Transmission Can Describe Many Synapses

The quantal nature of transmission was first demonstrated in the early 1950s by Bernard Katz and colleagues, who recorded from the frog neuromuscular junction. Spontaneous signals that were approximately 1/100 of the signal evoked by stimulating the presynaptic axon were observed to occur at random intervals. These spontaneous minis were of roughly constant amplitude, with a coefficient of variation of 30%. As discussed previously, a neuromuscular junction normally releases hundreds of quanta per impulse, but when extracellular Ca^{2+} was lowered to reduce transmitter release to very low levels, the evoked response fluctuated from trial to trial between preferred amplitudes, which coincided with integral multiples of the mini amplitude (Fig. 7.7).

We now understand this to arise from the fact that minis are the fusion of a single vesicle and that evoked responses can represent one or more of these same vesicles. Before the vesicle hypothesis was established, however, Katz and colleagues inferred that the mini is the quantal building block, variable numbers of which are released to make up the evoked signal. The relative numbers of trials resulting in 0, 1, 2, . . . quanta were well described by a Poisson distribution—a statistical distribution that arises in many instances where a random process operates. On the basis of this evidence, Katz and colleagues proposed the following model of transmission, which has gained wide acceptance and will be referred to as the standard Katz model.

- Arrival of an action potential at the presynaptic terminal briefly raises the probability of release of quanta of transmitter (i.e., the fusion of synaptic vesicles).
- Several quanta (synaptic vesicles) are available to be released, and every quantum gives roughly the same electrical signal in the postsynaptic cell. This is the

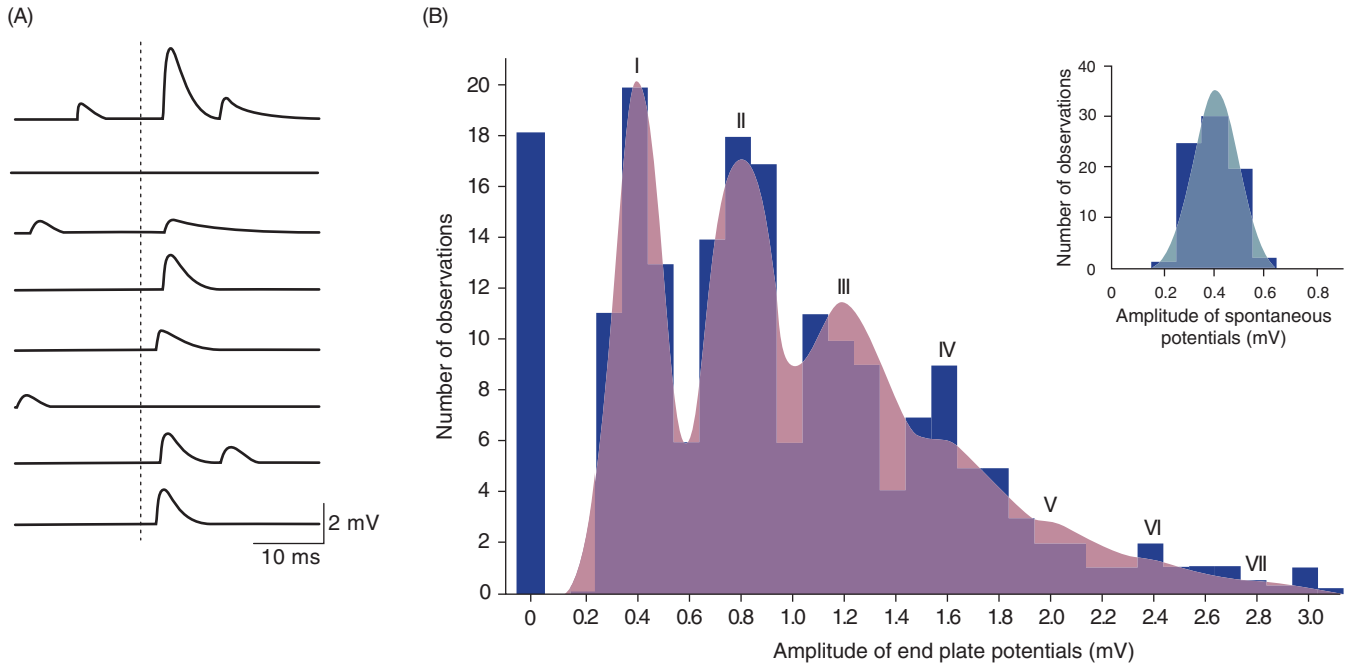


FIGURE 7.7 Quantal transmission at the neuromuscular junction. (A) Intracellular recordings from a rat muscle fiber in response to repeated presynaptic stimulation of the motor axon. Low extracellular [Ca^{2+}] and high [Mg^{2+}] restricted Ca^{2+} entry and so kept transmission to a very low level. The stimulus was given at the time marked by the dotted line. The size of the postsynaptic response fluctuated from trial to trial, with some trials giving failures of transmission. Spontaneous minis occurring in the background (e.g., those events that occur before the dotted line) had approximately the same amplitude as the smallest evoked responses, implying that they arose from the release of single quanta of acetylcholine. *From Liley (1956).* (B) Peak amplitudes of 200 evoked responses [end plate potentials (EPPs)] from a similar experiment, plotted as an amplitude histogram. Eighteen trials resulted in failures of transmission (indicated by the bar at 0 mV), and the rest gave EPPs whose amplitude tended to cluster at integral multiples of 0.4 mV. This coincides with the mean amplitude of the spontaneous minis, whose amplitude distribution is shown in the insert together with a Gaussian fit. Shading through the EPP histogram is a fit obtained by assuming a Poisson model of quantal release. Roman numerals indicate the number of quanta corresponding to each component in the distribution. *From Boyd and Martin (1956).*

quantal size or amplitude, Q, which sums linearly with all other quanta released.

- The average number of quanta released, m, is given by the product of n, the number of available quanta, and p, the average release probability: $m = np$. The average response to a stimulus is the product of the quantal size and the average number of quanta per stimulus, Qm or Qnp.
- The relative probability of observing 0, 1, 2, . . . , n quanta released is then given by a binomial distribution, with parameters n and p.

The value of Q has a clear underlying basis: the efficacy of an individual synaptic vesicle fusion in altering the conductance of the postsynaptic membrane. The binomial parameters, n and p, however, are formalisms whose underlying meanings are harder to pin down. n may correspond to the number of vesicles in a primed, fusion-ready state at the active zone or to discrete release sites competent for release, or even to the number of active zones. Correspondingly, p may represent the probability of fusion of any individual vesicle or the probability of fusion at a given release site or active zone. For some mechanisms, therefore, it is difficult to say whether n or p is altered. For example, if additional vesicles are recruited to release sites, that change may be manifest as a change in n because there are more releasable quanta or as a change in p because the probability of fusion at a given release site has increased because it is more likely to be charged with a vesicle.

These quantal parameters remain a useful description of the synapse, however, despite their obscurity on a mechanistic level. For example, it is often useful to classify synapses as having a high or low probability of release (p) to explain their behavior during trains of stimuli (see later). In a synapse with few releasable vesicles and a high p, a train is predicted to deplete the releasable pool rapidly and the strength of transmission declines, a phenomenon known as synaptic depression. Another synapse might have the same average quantal content, but accomplish this with a large releasable pool and a low probability of fusion for any individual vesicle in the pool. Such a synapse would be less prone to depression and indeed both types of synapse are encountered.

A major impetus for the accurate measurement of quantal parameters is that it may help determine the

locus of modulatory effects on synaptic transmission, or the site of action of a drug. A blocker of postsynaptic receptors, for example, will not alter the mini frequency or the number of quanta released, but will decrease the amplitudes of both spontaneous minis and the quantal components of evoked responses. A change in the loading of transmitter into a vesicle could cause a similar change. In contrast, a synaptic modulator that decreased Ca^{2+} channel opening or a blocker of those channels would decrease the probability of presynaptic release and be recognized as a decrease in quantal content, m. Many modulators of presynaptic transmission cause a parallel change in quantal content and in the frequency of spontaneous minis. This sort of analysis has been instrumental in understanding synaptic plasticity and in determining whether a given alteration occurs at the presynaptic terminal or the postsynaptic membrane.

The Standard Katz Model Does Not Always Apply

Incorrectly applying the Katz model and quantal analysis to synapses at which its assumptions do not hold has been a frequent source of error. Although a superb description of transmission at the neuromuscular junction, it is not uniformly applicable to central nervous system synapses, and the technical difficulties of isolating and accurately sampling the physiology of an individual synapse in the CNS can make the determination of quantal parameters exceptionally difficult.

In particular, five issues arise at CNS synapses with which the standard Katz model cannot deal adequately. These are *nonuniformity of quanta*, *nonuniformity of release sites*, *postsynaptic membranes that distort responses*, *saturation of receptors*, and *silent synapses*.

For quantal analysis to work, individual quanta should be roughly equivalent in size—that is, with the same amount of transmitter in each vesicle and with similar populations of receptors facing their release sites, as is true at the frog neuromuscular junction. But in the CNS, both vesicle content and receptor populations may differ widely, particularly where more than one synapse connects two cells. The resulting *nonuniformity of quanta* means that mini amplitudes will not have a normal distribution and will not sum to tidy peaks like those shown in Figure 7.7. Mini amplitude in the CNS can also be very hard to measure: the smallest events may be lost in the noise and the large variety of inputs converging on a postsynaptic cell makes it difficult to determine which responses derive from an individual presynaptic input. *Nonuniformity of release sites* can similarly prevent responses from falling in a simple binomial distribution. For example, the Ca channels at a nerve terminal may be of different isoforms or differentially phosphorylated and vesicles will therefore vary in their release probability depending on the channels closest to them. As evidence mounts that spontaneously fusing minis may arise from distinct vesicle pools and may activate different pools of receptors on postsynaptic cells, new questions arise as to whether the amplitude of spontaneous events can be used to infer the amplitude of the units underlying evoked responses.

Postsynaptic membranes that distort responses introduce another problem: the voltage-dependent properties of the postsynaptic cell and the particular electrical properties of dendritic spines can prevent vesicles from summing in a simple linear fashion. Equally problematic is the *saturation of receptors* on the postsynaptic membranes of many CNS synapses. A glutamatergic CNS synapse may have fewer than 100 receptors present, and if a single quantum activates most of these, it is impossible for quanta to sum linearly. Indeed, a major difference between some vertebrate CNS synapses and the neuromuscular junction is that the quantal amplitude often is determined not by the vesicle contents, but by the number of available receptors. In the extreme case of a developing synapse, the variation from impulse to impulse might depend as much on whether one, two, or three receptors were activated per vesicle fusion as whether one, two, or three vesicles fused.

Silent synapses represent an extreme form of the variability of postsynaptic responses and may be an important consideration for AMPA receptors at glutamatergic synapses (see Chapter 8). At these synapses, active zones may be secreting transmitter without producing a physiological response because no receptors are present in the postsynaptic membrane. These so-called silent synapses can spring to life when a set of receptors is inserted into the postsynaptic membrane in a concerted fashion. Because more active zones now are contributing to the detected physiological response, mini frequency will be higher and evoked release increased, although the quantal size will not have changed. Whereas the Katz model from the neuromuscular junction would consider these changes to be the hallmarks of a presynaptic change in release, the "un-silencing" of these CNS synapses is a postsynaptic event.

Thus, many of the requirements for the standard Katz model cannot realistically be expected to hold in all cases and the evidence that these simple probabilistic models are correct is far from compelling. However, despite the technical difficulties and theoretical caveats, the analysis of mini frequency, quantal size, quantal content, and synaptic probability has been the foundation of our description of synaptic transmission and the source of great insight into synaptic modulation.

It has been possible to augment this understanding with independent methods, including the ability to monitor vesicle fusion optically from the unloading of

the lipophilic dyes discussed earlier, and thereby to describe presynaptic activity without depending on the postsynaptic membrane to report the event. Another method of indirectly estimating release probability at glutamatergic synapses makes use of a blocker of NMDA receptors, MK-801, that will act only in the open state. From the rate of blockade during a train of stimuli, the probability of the channel having been opened by a given impulse can be inferred.

Summary

Quantal analysis has improved our understanding of biophysical and pharmacological mechanisms of transmission by describing normal transmission and by offering a tool to distinguish and characterize presynaptic and postsynaptic actions of drugs or endogenous agents that modify the strength of a synapse. Although these methods must be applied with caution, they yield a unique insight into the mechanisms of synaptic plasticity. The nature of transmission at CNS synapses can be quite different from what is encountered at the classical synapse of the neuromuscular junction.

SHORT-TERM SYNAPTIC PLASTICITY

Chemical synapses are not static transmitters of information. Their effectiveness waxes and wanes, depending on the frequency of stimulation and the history of prior activity. Because information processing in the nervous system in not conducted by individual action potentials at widely spaced intervals, this plasticity is of the utmost importance to the function of the synapse within its physiological circuit. At most synapses, repetitive high-frequency stimulation (called a *tetanus*) initially is dominated by a growth in successive PSP amplitudes, called synaptic *facilitation* (Fig. 7.8). This process builds to a steady state within about 1 s and decays equally rapidly when stimulation stops. Decay is measured by single test stimuli given at various intervals after a conditioning train. Facilitation does not require a train to be observed; at many synapses it can be seen after a single action potential. Thus, when two stimuli are given with very close spacing, the second one can be as much as twice the amplitude of the first. This is called *paired-pulse facilitation* (Fig. 7.8). At most synapses, a slower phase of increase in efficacy, which has a characteristic time constant of several seconds (called *augmentation*), succeeds facilitation. Finally, with prolonged stimulation, some synapses display a third phase of growth in PSP amplitude that lasts minutes (called *potentiation*). Potentiation often is visible in isolation only long after a tetanus and is thus called *posttetanic potentiation* (PTP).

Often, a phase of decreasing transmission, called synaptic *depression*, is superimposed on these processes (Fig. 7.8). Synaptic depression leads to a dip in transmission during repetitive stimulation, which often tends to overlap and obscure the augmentation and potentiation phases. At some synapses, it is manifest after only a single action potential and gives rise to *paired-pulse depression*.

Additional forms of use-dependent plasticity, called long-term potentiation and long-term depression (LTP and LTD), have received considerable attention as likely correlates of some forms of learning and are discussed elsewhere (Chapter 47). Most frequently, at synapses in which a quantal analysis has been done, all these forms of synaptic plasticity (except some forms of cortical LTP) are due to changes in the number of quanta released by action potentials.

The propensity to facilitation or depression can be quite different at different types of synapse and thus contributes to the diversity of synaptic communication and the flexibility of information processing by synapses. At a given synapse, the frequency and duration of a train can alter whether facilitation or depression is most in evidence. In a physiologically realistic setting of irregularly spaced stimuli, all these phenomena may be simultaneously at play and one or another may predominate for any given response, depending on the preceding history of stimulation. Thus, responses in the postsynaptic cell will be anything but a constant response to each action potential that comes along.

Synaptic Depression May Arise from Depletion of Readily Releasable Vesicles, Autoinhibition, or Receptor Desensitization

The rate at which synaptic depression develops usually depends on stimulation frequency. At many synapses, depression is relieved when transmission is reduced by lowering Ca^{2+} or raising Mg^{2+} in the medium—when little or no release occurs, there is little depression of the subsequent response. These characteristics are consistent with depression being due to the depletion of a readily releasable store of docked or nearly docked vesicles and recovery being due to their replenishment from a nearby supply (Fig. 7.9). The parameters of such a depletion model can be estimated from the rate of recovery from depression, which gives the rate of refilling the releasable store, and the fractional drop in PSPs given at short intervals, which gives the fraction of the releasable store liberated by each action potential.

As mentioned previously, synapses with a high probability of release from a relatively small pool of vesicles

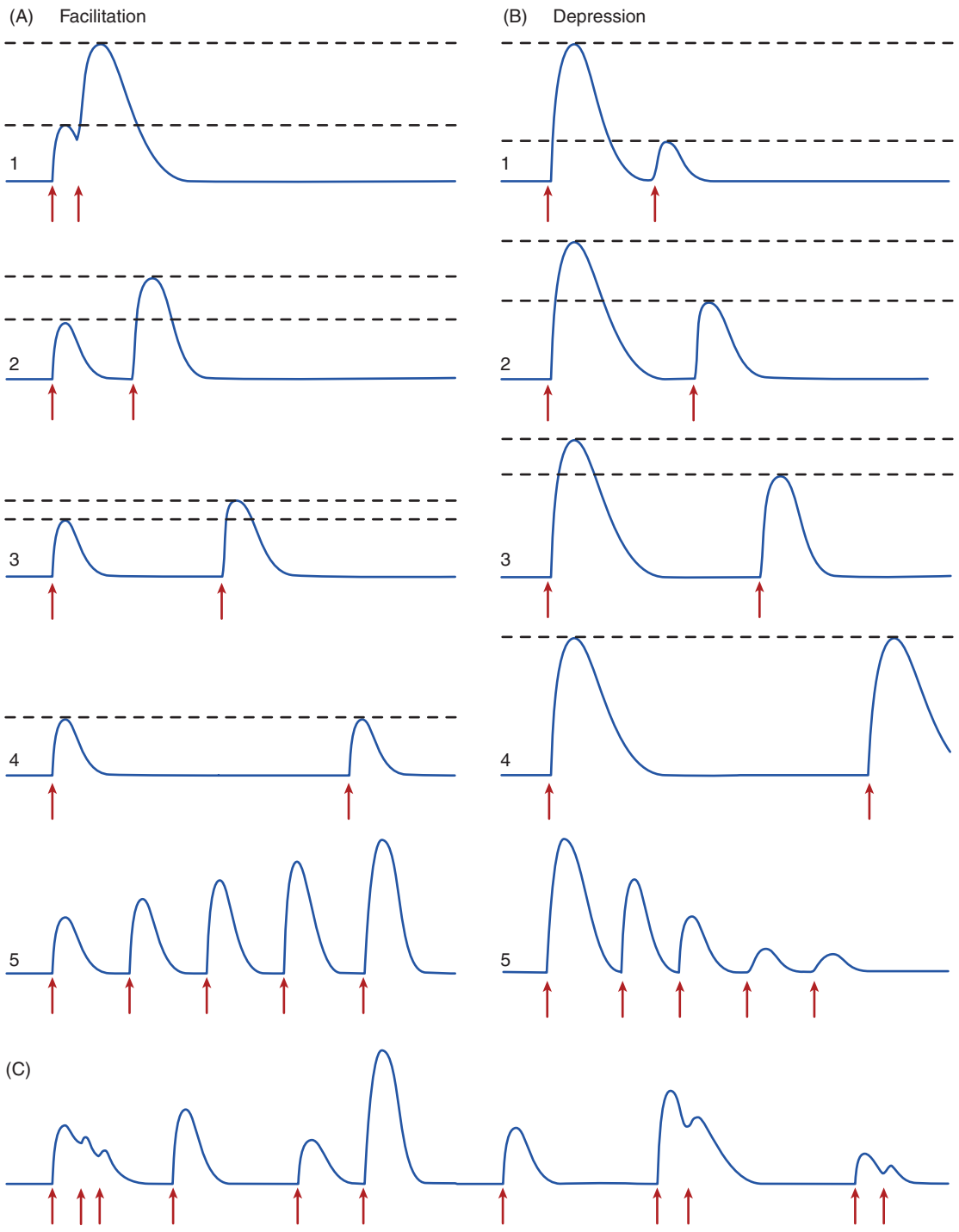


FIGURE 7.8 Short-term synaptic plasticity. (A) An idealized synapse in which facilitation predominates. Two stimuli (paired pulses) are given to the presynaptic nerve in each line and postsynaptic voltage responses are shown. Closely spaced stimuli (top line) facilitate strongly but as the interval between stimuli increases, the degree of facilitation diminishes. The fifth line illustrates how facilitation can accumulate during a train of stimuli. (B) An idealized synapse in which synaptic depression predominates. (C) At most synapses, both depression and facilitation occur and, depending on the nature of the individual synapse and the time course of facilitation, depression, augmentation, and potentiation, any of these phenomena may predominate at a given moment. Thus, in a realistic pattern of activity, with unevenly spaced action potentials arriving at a terminal, the strength of transmission can depend greatly on the previous history of the synapse.

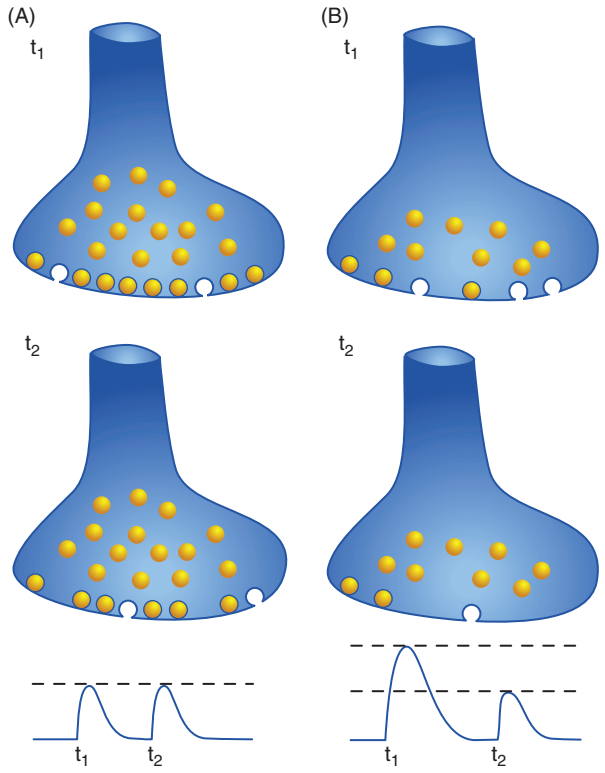


FIGURE 7.9 Depletion model of synaptic depression depicting the action of two stimuli (t_1 and t_2) that arrive so closely spaced that no vesicles have been recruited to the releasable pool to replace those that fused in response to the first stimulus. (A) A synapse in which the releasable pool (n) is large and the probability of fusion for a given vesicle (p) is low shows little or no depression. Because only a small fraction of the docked, releasable vesicles fused in response to the first stimulus, the second response is comparable in amplitude. (B) A synapse in which the releasable pool (n) is small but the probability of fusion (p) is high is likely to be strongly depressing. Vesicles released by the first impulse deplete half the fusion-competent pool, leaving n at t_2 equal to half its value at t_1. If p has not changed at t_2, the second stimulus will release half as many quanta.

will be the most prone to this form of depression (Fig. 7.9B). Recovery from this form of depression appears to be influenced by cytosolic Ca^{2+} as well, consistent with a hypothetical ability of elevated Ca^{2+} in the cytosol to enhance the recruitment of reserve vesicles to release sites.

Vesicle depletion cannot account for all forms of synaptic depression. At some synapses, depression is due to an inhibitory action of released transmitter on presynaptic receptors called autoreceptors. For example, in rat hippocampal cortex, the depression of GABA responses is blocked by antagonists of presynaptic $GABA_B$ receptors. Presumably, GABA acts to hyperpolarize nerve terminals and block transmitter release. At other synapses, desensitization of postsynaptic receptors contributes heavily to the observed depression; although the same amount of transmitter is released with each impulse, the postsynaptic receptors have a diminished ability to respond (Chapter 8).

Facilitation, Augmentation, and Potentiation Are Due to Effects of Residual Ca^{2+}

With few exceptions, all the phases of increased short-term plasticity are Ca^{2+}-dependent in the sense that little or no facilitation, augmentation, or potentiation is generated by stimulation in Ca^{2+}-free medium. Originally, these phases of increased transmission were thought to be due to the effect of residual Ca^{2+} remaining in active zones after presynaptic activity and summating with Ca^{2+} influx during subsequent action potentials to generate slightly higher peaks of Ca^{2+} (Katz & Miledi, 1968). Due to the highly nonlinear dependence of transmitter release on Ca^{2+}, a small residual amount could activate a substantial increase in phasic transmitter release during an action potential, while having only a small effect on mini frequency itself. However, as this relationship was examined more closely and with Ca^{2+} indicators to actually measure the residual intracellular Ca^{2+}, this hypothesis became unlikely. These findings led to the proposal that, in addition to summating with peak Ca^{2+} transients to drive fusion, Ca^{2+} acts to increase transmission at one or more targets distinct from the sites triggering exocytosis. Unlike the triggering site, the facilitation site for Ca^{2+} binding should be a high-affinity site that can be influenced by the relatively low levels of residual Ca^{2+}. Other second messenger systems and kinases may also be involved.

Potentiation lasts for a long period after a strong tetanus because residual Ca^{2+} is present for the duration of PTP, presumably due to overloading of the processes responsible for removing excess Ca^{2+} from neurons. These processes include Ca^{2+} extrusion pumps, such as plasma membrane ATPase and Na^{2+}-Ca^{2+} exchange, and Ca^{2+} uptake into organelles such as endoplasmic reticulum and mitochondria.

Summary

Short-term synaptic plasticity allows synaptic strength to be modulated as a function of prior activity and can cause large changes in responses during physiologically relevant patterns of stimulation. Synapses may show a decline in transmission (depression) or an increase in synaptic efficacy with time constants ranging from seconds (facilitation and augmentation) to minutes (potentiation or PTP) to hours (LTP); many synapses show a mixture of several of these phases. Synaptic depression may be due to depletion of a readily releasable supply of vesicles, to the inhibitory action of transmitter on presynaptic autoreceptors, or to the desensitization of postsynaptic receptors. Depression makes synapses selectively responsive to brief stimuli or to changes in level of activity. Frequency-dependent increases in synaptic efficacy are due to the effects of residual presynaptic Ca^{2+} acting to modulate the release process, probably through

a high-affinity binding site. These frequency-dependent increases in synaptic efficacy allow synapses to distinguish significant signals from noise and to respond to selected patterns of activity.

References

Bennett, M. K., & Scheller, R. H. (1993). The molecular machinery for secretion is conserved from yeast to neurons. *Proceedings of the National Academy of Sciences of the United States of America, 90*(7), 2559–2563.

Boyd, I. A., & Martin, A. R. (1956). Spontaneous subthreshold activity at mammalian neuromuscular junctions. *Journal of Physiology, 132,* 74–91.

Gonzalez, L., & Scheller, R. H. (1999). Regulation of membrane trafficking: Structural insights from a Rab/effector complex. *Cell, 96*(6), 755–758.

Harlow, M. L., Ress, D., Stoschek, A., Marshall, R. M., & McMahan, U. J. (2001). The architecture of active zone material at the frog's neuromuscular junction. *Nature, 409*(6819), 479–484.

Heuser, J. E. (1977). Synaptic vesicle exocytosis revealed in quick-frozen frog neuromuscular junctions treated with 4-aminopyridine and given a single electrical shock. *Society for Neuroscience Symposium, 2,* 215–239.

Heuser, J. E., Reese, T. S., Dennis, M. J., Jan, Y., Jan, L., & Evans, L. (1979). Synaptic vesicle exocytosis captured by quick freezing and correlated with quantal transmitter release. *Journal of Cell Biology, 81*(2), 275–300.

Katz, B. (1969). The release of neural transmitter substances. Springfield, IL: Thomas.

Katz, B., & Miledi, R. (1968). The role of calcium in neuromuscular facilitation. *Journal of Physiology, 195*(2), 481–492.

Liley, A. W. (1956). The quantal components of the mammalian end-plate potential. *Journal of Physiology, 133,* 560–573.

Rizo, J., & Rosenmund, C. (2008). Synaptic vesicle fusion. *Nature Structural & Molecular Biology, 15*(7), 665–674.

Schneggenburger, R., & Neher, E. (2000). Intracellular calcium dependence of transmitter release rates at a fast central synapse. *Nature, 406*(6798), 889–893.

Sollner, T., Bennett, M. K., Whiteheart, S. W., Scheller, R. H., & Rothman, J. E. (1993a). A protein assembly-disassembly pathway in vitro that may correspond to sequential steps of synaptic vesicle docking, activation, and fusion. *Cell, 75*(3), 409–418.

Sollner, T., Whiteheart, S. W., Brunner, M., Erdjument-Bromage, H., Geromanos, S., Tempst, P., & Rothman, J. E. (1993b). SNAP receptor implicated in vesicle targeting and fusion. *Nature, 362*(6418), 318–324.

Sutton, R. B., Fasshauer, D., Jahn, R., & Brunger, A. T. (1998). Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 A resolution. *Nature, 395*(6700), 347–353.

Takamori, S., Holt, M., Stenius, K., Lemke, E. A., Grønborg, M., Riedel, D., … Jahn, R. (2006). Molecular anatomy of a trafficking organelle. *Cell, 127*(4), 831–846.

Yao, J., Gaffaney, J. D., Kwon, S. E., & Chapman, E. R. (2011). Doc2 is a Ca^{2+} sensor required for asynchronous neurotransmitter release. *Cell 147*(3), 666–677.

Suggested Readings

Dittman, J. S., & Regehr, W. G. (1998). Calcium dependence and recovery kinetics of presynaptic depression at the climbing fiber to Purkinje cell synapse. *Journal of Neuroscience, 18*(16), 6147–6162.

Ferguson, S. M., Raimondi, A., Paradise, S., Shen, H., Mesaki, K., Ferguson, A., … De Camilli, P. (2009). Coordinated actions of actin and BAR proteins upstream of dynamin at endocytic clathrin-coated pits. *Developmental Cell, 17*(6), 811–822.

Kochubey, O., Lou, X., & Schneggenburger, R. (2011). Regulation of transmitter release by Ca^{2+} and synaptotagmin: Insights from a large CNS synapse. *Trends in Neuroscience, 34*(5), 237–246.

Raimondi, A., Ferguson, S. M., Lou, X., Armbruster, M., Paradise, S., Giovedi, S., … De Camilli, P. (2011). Overlapping role of dynamin isoforms in synaptic vesicle endocytosis. *Neuron, 70*(6), 1100–1114.

Ramirez, D. M., & Kavalali, E. T. (2011). Differential regulation of spontaneous and evoked neurotransmitter release at central synapses. *Current Opinion in Neurobiology, 21*(2), 275–282.

Ryan, T. A., Li, L., Chin, L. S., Greengard, P., & Smith, S. J. (1996). Synaptic vesicle recycling in synapsin I knock-out mice. *Journal of Cell Biology, 134*(5), 1219–1227.

Ryan, T. A., Reuter, H., Wendland, B., Schweizer, F. E., Tsien, R. W., & Smith, S. J. (1993). The kinetics of synaptic vesicle recycling measured at single presynaptic boutons. *Neuron, 11*(4), 713–724.

Schiavo, G., Rossetto, O., Tonello, F., & Montecucco, C. (1995). Intracellular targets and metalloprotease activity of tetanus and botulism neurotoxins. *Current Topics in Microbiology and Immunology, 195,* 257–274.

Wierda, K. D., Toonen, R. F., de Wit, H., Brussaard, A. B., & Verhage, M. (2007). Interdependence of PKC-dependent and PKC-independent pathways for synaptic plasticity. *Neuron, 54*(2), 275–290.

Zucker, R. S. (1989). Short-term synaptic plasticity. *Annual Review of Neuroscience, 12,* 13–31.

Thomas L. Schwarz

This page intentionally left blank

CHAPTER

8

Neurotransmitter Receptors

Chemical synaptic transmission plays a fundamental role in neuron-to-neuron and neuron-to-muscle communication and is mediated by receptors embedded in the plasma membrane. The cellular response is determined by which specific receptor(s) the transmitter binds, and the response can be either excitatory or inhibitory. This important feature of chemical synaptic transmission is termed *sign inversion* because excitation of one neuron in the circuit can cause inhibition of the follower neuron. The response magnitude is ultimately determined by receptor number, the "state" of the receptors, and the amount of transmitter released. Finally, the temporal and spatial summation of information conveyed by the activation of multiple receptors is the determinant of whether that neuron will fire an action potential or the muscle will contract. As one can see, there is remarkable flexibility and diversity in molding the response to neurotransmitter by constructing a synapse with the desired receptor types.

Two broad classifications exist for receptors: ionotropic and G-protein coupled (GPCR). An *ionotropic receptor* is a relatively large, multisubunit complex typically composed of four or five individual proteins that combine to form an ion channel through the membrane (Fig. 8.1A). In the absence of neurotransmitter, these ion channels exist in a closed state and are largely impermeable to ions. Neurotransmitter binding induces rapid conformational changes that open the channel, permitting ions to flow down their electrochemical gradients. Changes in membrane current resulting from ligand binding to ionotropic receptors generally are measured on a millisecond time scale. The ion flow ceases when the transmitter dissociates from the receptor or when the receptor becomes desensitized. In contrast, a GPCR is typically composed of a single polypeptide (Fig. 8.1B), although dimeric forms exist, and exerts its effects not through the direct opening of an ion channel but through binding to and activating GTP-binding proteins (G-proteins). Transmitters that activate GPCRs typically produce responses of slower onset and longer duration (from tenths of seconds to potentially hours) due to the series of enzymatic steps involved. However, faster responses can occur through actions of activated G-proteins, or their associated subunits, directly on ion channels.

IONOTROPIC RECEPTORS

All ionotropic receptors are membrane-bound protein complexes that form an ion-permeable pore in the membrane. By comparing the amino acid sequence of cloned ionotropic receptors, one can deduce that they are similar in overall structure, although two independent ancestral genes gave rise to two distinct families (Fig. 8.2). One family includes the nicotinic acetylcholine (Ach) receptor (nAChR), the γ-aminobutyric acid (GABA) receptor ($GABA_A$), the glycine receptor, and one subclass of serotonin (5-HT) receptor, the 5HT3 receptor. The other family comprises the set of ionotropic glutamate receptors (Hollmann & Heinemann, 1994; Traynelis et al., 2010).

The nAChR is a Heteromeric Protein Complex with Distinct Architecture

The nAChR is so named because the plant alkaloid nicotine can bind to the ACh-binding site and activate the receptor. The nAChR purified from the electric organ of *Torpedo* ray is composed of five subunits (Fig. 8.1A) and has a native molecular mass of approximately 290 kDa. The subunits are designated α, β, γ, and δ, and each receptor complex contains two copies of the α subunit. The subunits are homologous membrane-bound proteins that assemble in the bilayer to form a ring enclosing a

© 2013 Elsevier Inc. All rights reserved.

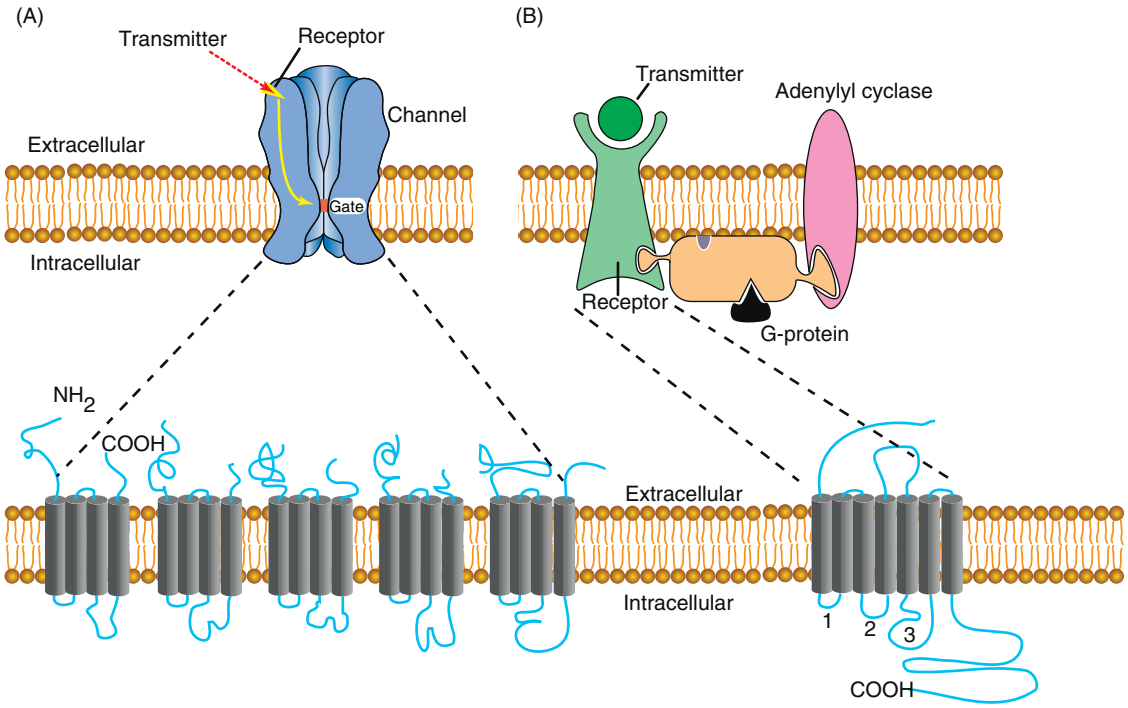


FIGURE 8.1 A comparison of the general structural features of ionotropic and G-protein couples receptors. (A) *Ionotropic receptors* bind transmitter, and this binding translates directly into the opening of the ion channel through a series of conformational changes. Ionotropic receptors are composed of multiple subunits. The five subunits that together form the functional nAChR are shown. Note that each of the nAChR subunits wraps back and forth through the membrane four times and that the mature receptor is composed of five subunits. (B) *G-protein coupled receptors* bind transmitter and, through a series of conformational changes, bind to G-proteins and activate them. G-proteins then activate enzymes such as adenylyl cyclase to produce cAMP. Through the activation of cAMP-dependent protein kinase, ion channels become phosphorylated, which affects their gating properties. GPCRs are single subunits or dimers. They contain seven transmembrane-spanning segments, with the cytoplasmic loops formed between the segments providing the points of interactions for coupling to G-proteins.

central pore. The extracellular domain of each subunit together forms a funnel-shaped opening that extends approximately 100 Å outward from the outer leaflet of the plasma membrane. The funnel at the outer portion of the receptor has an inside diameter of 20–25 Å. The funnel shape is thought to concentrate and force ions to interact with amino acids in the limited space of the pore without producing a major barrier to diffusion. This funnel narrows near the center of the lipid bilayer to form the domain of the receptor that determines the opened or closed state of the ion pore (Fig. 8.1). The intracellular domain of the receptor forms short exits for ions traveling into the cell and an entrance for ions traveling out of the cell. The intracellular domain also establishes the association of the receptor with other intracellular proteins that determine the subcellular localization of the nAChR.

Each nAChR Subunit Has Multiple Membrane-Spanning Segments

The primary structure of each nAChR subunit was obtained by the efforts of Shosaka Numa and colleagues. The deduced amino acid sequence from cloned mRNAs indicates that nAChR subunits range in size from 40 to 65 kDa. Each subunit consists of four transmembrane (TM)-spanning segments referred to as TM1–TM4 (Fig. 8.3A). Each segment is composed mainly of hydrophobic amino acids that stabilize the domain within the hydrophobic environment of the lipid membrane. The four TM domains are arranged in an antiparallel fashion, wrapping back and forth through the membrane. The N terminus of each subunit extends into the extracellular space, as does the loop connecting TM2 and TM3, as well as the C terminus (Fig. 8.3A). Amino acids linking TM1 and TM2 and those linking TM3 and TM4 form short loops that extend into the cytoplasm.

Structure of the Channel Pore Determines Ion Selectivity and Current Flow

In the model shown in Figure 8.3B and 8.3C, each subunit of the nAChR can be seen to contribute one cylindrical component (representing a membrane-spanning segment) that presents itself to a central cavity that forms the ion channel through the center of the complex. The membrane-spanning segments that line the pore are

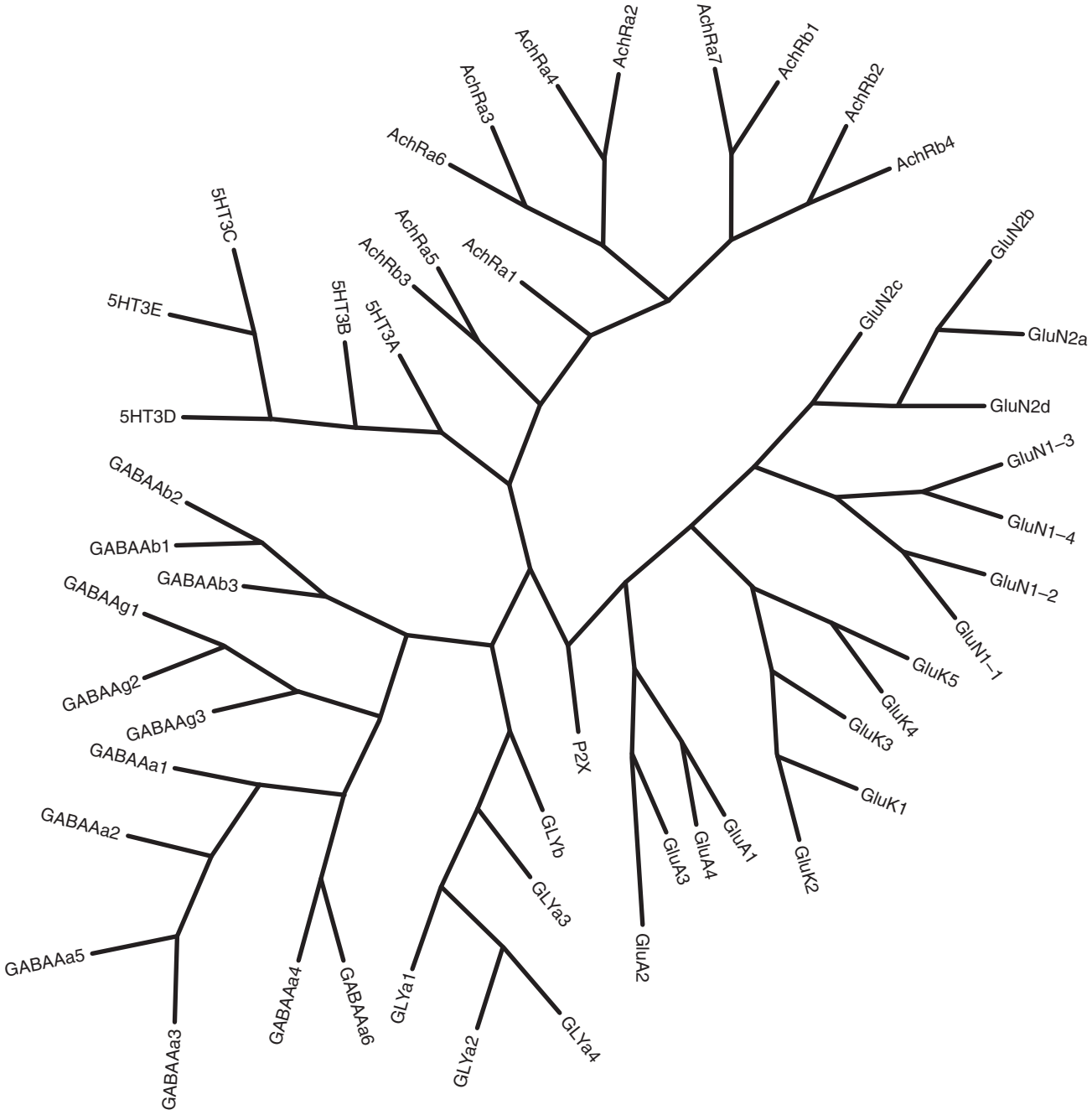


FIGURE 8.2 Evolutionary relationships of the ionotropic receptor family. The tree was constructed by aligning the protein sequences from each receptor family with the ClustalW program. Based on the alignment, the phylogenetic relationship was inferred with the maximum parsimony method and the tree was constructed using the Phylogeny Inference Package v 3.6 (distributed by J. Felsenstein, Department of Genome Sciences, University of Washington, Seattle, WA). Dr. Yin Liu (Department of Neurobiology and Anatomy, University of Texas Health Science Center-Houston, Houston, TX) kindly provided the phylogenetic tree and figure.

the five TM2 regions, one contributed by each subunit. The amino acids that compose the TM2 segment are arranged in such a way that three rings of negatively charged amino acids are oriented toward the central pore of the channel (Fig. 8.3A). These rings of negative charge appear to contribute to the selectivity filter of the channel, increasing the probability that only cations can pass through the pore. The nAChR is permeable to most cations, such as Na^+, K^+, and Ca^{2+}, although monovalent cations are preferred. The restricted physical dimensions of the pore—9–10 Å in the open state—contribute greatly to the selectivity for particular ions (Unwin, 1995, 2005). A coarse filtering that also influences selectivity appears to be a shielding effect produced by other negatively

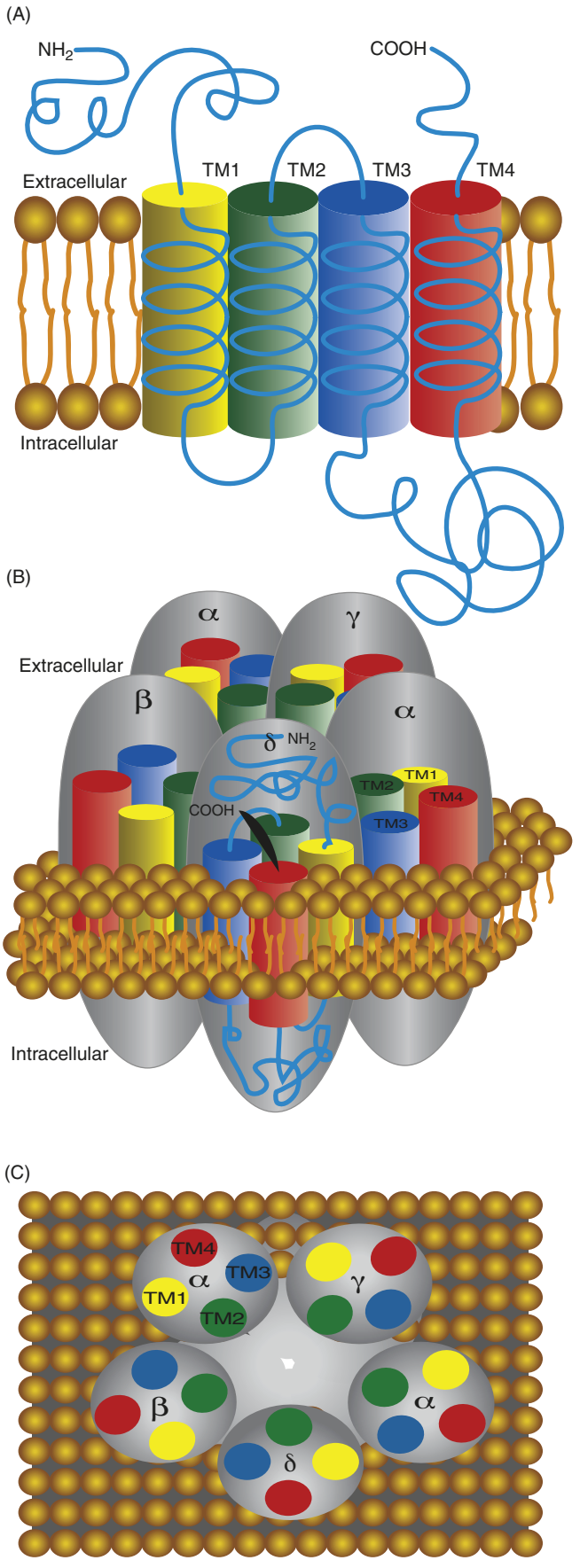


FIGURE 8.3 (A) Diagram highlighting the orientation of membrane-spanning segments of one subunit of the nAChR. The amino and carboxy termini extend in the extracellular space. The four membrane-spanning segments are designated TM1–TM4. Each forms an α helix as it traverses the membrane. (B) Side view of the five subunits in their approximate positions within the receptor complex. There are two α subunits present in each nAChR. (C) Top view of all five subunits highlighting the relative positions of their membrane-spanning segments, TM1–TM4, and the position of TM2 that lines the channel pore.

charged amino acids surrounding the outer channel region of the receptor. Collectively, these physical characteristics of the pore—together with the electrochemical gradient across the plasma membrane—determine the possibility and rate of ion movements. Thus, when the pore of the nAChR opens, anions are restricted from movement across the membrane, while positively charged cations move down their respective electrochemical gradients, resulting in an influx of Na^+ and Ca^{2+} and a small efflux of K^+.

Opening of the nAChR Occurs through Concerted Conformational Changes Induced by Binding of Two ACh Molecules

Each receptor complex has two ACh-binding sites that reside in the extracellular domain and are formed for the most part by six amino acids in α subunits; however, amino acids in both γ and δ subunits also contribute to binding. The two binding sites are not equivalent because of the asymmetry of the receptor produced by the different neighboring subunits (either γ or δ) adjacent to the α subunits. Significant cooperativity also exists within the receptor molecule, and so binding of the first molecule of ACh enhances binding of the second. When nAChR binds two molecules of ACh, the channel opens almost instantaneously (time constants for opening are approximately 20 μs), thus permitting the passage of ions. A model developed from electron micrographic reconstructions of the nicotine-bound form of the nAChR indicates that the closed-to-open transition is associated with a rotation of the TM2 segments (Fig. 8.4; Unwin, 1995, 2005). The TM2 segments are helical and exhibit a kink in their structure that forces a Leu residue from each segment into a tight ring that effectively blocks the flow of ions through the central pore of the receptor. When the TM2 segments rotate because of ACh binding, the kinks also rotate, relaxing the constriction formed by the Leu ring, and ions can then permeate through the pore. The rotation also orients a series of Ser and Thr residues (amino acids with a polar character) into the central area of the pore (compare Figs. 8.4B to 8.4C), which facilitates the permeation of water-solvated cations.

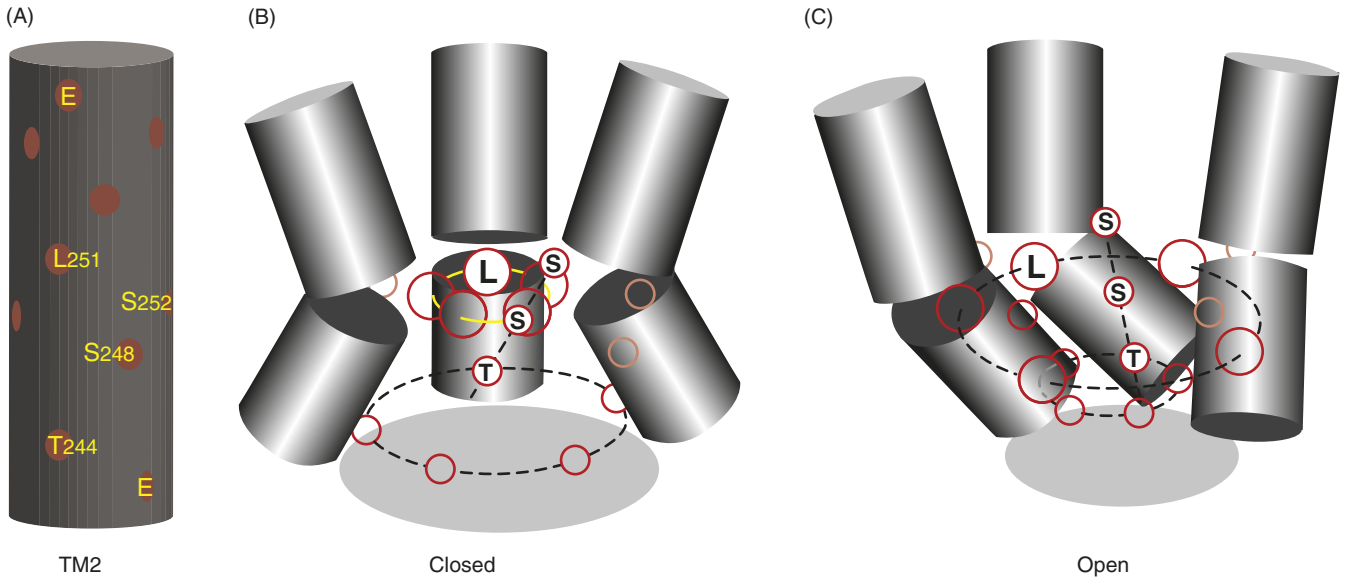


FIGURE 8.4 (A) Relative positions of amino acids in the TM2 segment of one of the nAChR *a* subunits modeled as an *a* helix. Glutamate residues (E) that form parts of the negatively charged rings for ion selectivity are shown at the top and bottom of the helix. (B) Arrangement of three of the five TM2 segments of the nAChR modeled with the receptor in the closed (ACh-free) configuration. In the closed configuration, leucine (L) residues form a right ring in the center of the pore that blocks ion permeation. (C) Arrangement of the three TM2 segments after ACh binds to the receptor. In the open configuration, construction formed by the ring of leucine (L) residues opens as the helices twist about their axes. Note that polar serine (S) and threonine (T) residues align when ACh binds, which apparently help the water-solvated ions travel though the pore. *Adapted from Unwin (1995).*

The Muscle Form of the nAChR is Very Similar to the nAChR from *Torpedo*

nAChRs at the neuromuscular junction are a concentrated collection of homogeneous receptors having a structure similar to that of the *Torpedo* electric organ. This similarity is not surprising because the electric organ is a specialized form of muscle tissue. The adult form of the muscle receptor has the pentameric structure $\alpha_2\beta\varepsilon\delta$ and is highly similar in structure to the nAChR from *Torpedo*. An embryonic form of the muscle receptor has an analogous structure, except that the ε subunit is replaced by a unique γ subunit. The embryonic and adult subunits of both mouse and bovine muscle receptors have been cloned and expressed and the receptors differ in both channel kinetics and channel conductance. These differences in channel properties that occur due to subunit swapping appear to be necessary for the proper function of the nAChRs during the transition from developing to mature neuromuscular junction synapses.

The nAChR Has Well-Ordered Assembly, is Post-translationally Modified, and is Concentrated and Anchored in the Postsynaptic Membrane

The pathway of nAChR assembly in muscle is a tightly regulated process. The five subunits of the nAChR have the potential to assemble randomly into 208 different combinations. Nevertheless, in vertebrate muscle, only one of these configurations ($\alpha_2\beta\varepsilon\delta$) is typically found in mature tissue, indicating a very high degree of coordinated assembly and, ultimately, little structural variability. The well-ordered assembly takes place within the endoplasmic reticulum, and during intracellular maturation, each subunit is glycosylated. Two highly conserved disulfide bonds in the *N*-terminal extracellular domain are essential for efficient assembly of the mature receptor. The first is between two adjacent Cys residues that reside very close to the ACh-binding site on the receptor. The second bond is between two Cys residues 15 amino acids apart, forming a loop that is spatially adjacent to the TM domains of the native receptor. This "Cys loop" is a structural signature of this family of receptors and is also found in the $GABA_A$ receptor, the glycine receptor, the 5HT3 receptor, and two members of the purinergic receptors.

Many ionotropic receptors, such as the nAChR, are phosphorylated, although the functional significance of the phosphorylation is not always evident. The nAChR is phosphorylated by at least three protein kinases. cAMP-dependent protein kinase (PKA) phosphorylates the γ and δ subunits, Ca^{2+}/phospholipid-dependent protein kinase (PKC) phosphorylates the δ subunit, and an unidentified tyrosine kinase phosphorylates the β, γ, and δ subunits. The phosphorylation sites are all found in the intracellular loop between TM3 and TM4

membrane-spanning segments. Phosphorylation by these three protein kinases increases the rate of desensitization of the receptor. Desensitization of receptors is common, and this process limits the amount of ion flux through a receptor by producing transitions into a closed state (one that does not permit ion flow) in the continued presence of neurotransmitter. For the nAChR, the rate of desensitization has a time constant of approximately 50–100 ms.

nAChRs are highly concentrated at the neuromuscular junction that ensures rapid and reliable communication between the presynaptic motor neuron and postsynaptic muscle cell to induce contraction (Fig. 8.5). The clustering of nAChRs begins during development, and is mediated in part by the release of agrin from the motor neuron that binds to a tyrosine kinase in the postsynaptic membrane called MuSK (Willmann & Fuhrer, 2002). MuSK activation recruits members of the Src family of kinases that phosphorylates the nAChR, leading to increased interactions with a multitude of postsynaptic signaling complexes ultimately leading to the recruitment of rapsyn. Rapsyn is a multidomain protein that binds directly to cytoplasmic tails of the nAChR subunits along with other proteins, including itself. Rapsyn thus forms the critical scaffolding function that clusters nAChRs at the NMJ. In addition to the nAChR, rapsyn binds to other important muscle proteins including MuSK and the dystrophin/utrophin glycoprotein complex (dystrophin is the molecule that when mutated leads to certain types of muscular dystrophy). Activity of the nAChR itself is also critical for its stabilization at the neuromuscular junction. For example, the lifetime of the nAChR at the junction decreases from 14 d to less than a day following block of the receptor with α-bungarotoxin. The interplay of signals coming from the presynaptic motor neuron is integrated by the postsynaptic receptors to concentrate and stabilize AChRs at the neuromuscular junction.

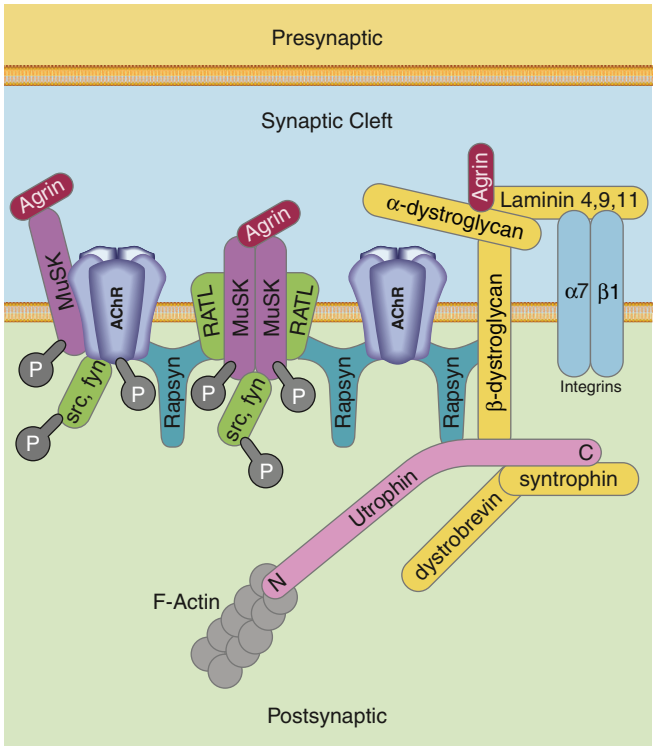


FIGURE 8.5 Diagram of nAChR clustering at the neuromuscular junction. Rapysn is a major anchoring protein at the neuromuscular junction that binds to itself and to the nAChR that concentrates and stabilizes nAChRs. The development and stabilization of the neuromuscular junction is mediated by a number of signaling cascades, only a few of which are shown. For example, agrin released from the presynaptic motor neuron binds to a number of proteins associated with the postsynaptic membrane including the tyrosine kinase MuSK (muscle specific kinase). MuSK activation by agrin recruits and activates the soluble tyrosine kinases src and fyn that further modify a number of proteins. RATL (rapsyn associated linker protein) is a membrane-bound protein that binds to both MuSK and to rapsyn to anchor MuSK at the neuromuscular junction. Agrin also interacts with the dystroglycans that make up the dystrophin complex important for the maintenance of the neuromuscular junction. Rapsyn also binds to the utrophin complex that anchors the overlying protein complex to the actin cytoskeleton. *Adapted from Willmann and Fuhrer (2002).*

Neuronal nAChRs Contain Two Types of Subunits and Their Assembly Leads to Functional Diversity

Neuronal nAChRs are similar yet distinct in structure to the *Torpedo* isoform of the receptor. For example, the neuronal nAChR is composed of only two types of subunits, α and β, which combine to produce the functional receptor, and the majority of these receptors do not bind to α-bungarotoxin. At least ten different α subtypes (α1 being the muscle α subunit) have been identified (see Fig. 8.2 for comparisons), and some are species-specific (α8 is found only in chicken and α9 is found only in rat). Four different β subtypes (β1 being the muscle β subunit) have been identified. The neuronal β subunits are distantly related to the muscle β1 subunit and sometimes are referred to simply as non-α subunits. All the α and β genes encode proteins with four transmembrane-spanning segments, and although the physical structure of this receptor family has not been well characterized, it appears that each functional receptor is a pentameric assembly.

Neuronal nAChRs have diverse functions and are the receptors presumed to be responsible for the psychophysical effects of nicotine addiction. One major function of nAChRs in the brain is to modulate excitatory synaptic transmission through a presynaptic action. The diversity in function of nAChRs can be related to the

heterogeneous structure contributed by the thousands of possible combinations between the different α and β subunits. Functional neuronal nAChRs can be assembled from a single subunit (e.g., $\alpha 7$, $\alpha 8$, or $\alpha 9$), or a single type of α subunit can be assembled with multiple types of β subunits (e.g., $\alpha 3$ with $\beta 2$ or $\beta 4$ or both) and vice versa. The mammalian high affinity nicotine-binding receptor appears to be formed by the $\alpha 4$ and $\beta 2$ subunits. These subunit-mixing possibilities produce a staggering array of potential receptor molecules, each with distinct properties, including differences in single-channel kinetics and rates of desensitization. It is also now established that subunit composition plays important roles in targeting the receptors to different intracellular locations.

Neuronal nAChRs exhibit a range of single-channel conductance between 5 and 50 pS, depending on the tissue or the specific subunits expressed. All the neuronal nAChRs are cation-permeable channels that, in addition to permitting the influx of Na^+ and the efflux of K^+, permit an influx of Ca^{2+}. The Ca^{2+} permeability for neuronal nAChR is greater than that for the muscle nAChR and is variable among the different neuronal receptor subtypes. Indeed, some receptors have very high Ca^{2+}/Na^+ permeability ratios; for example, $\alpha 7$ nAChRs exhibit a Ca^{2+}/Na^+ permeability ratio of nearly 20, whereas other neuronal isoforms exhibit Ca^{2+}/Na^+ permeability ratios of about 1.0–1.5. The Ca^{2+} permeability of the $\alpha 7$ nAChR can be eliminated by the mutation of a single amino acid residue in TM2 (Glu-237 for Ala) without significantly affecting other aspects of the receptor. This key Glu residue presumably lies at a critical position within the pore of the receptor and enhances the passage of Ca^{2+} ions through an interaction with its negatively charged side chain. Activation of $\alpha 7$ receptors through the binding of ACh therefore could produce a significant increase in the level of intracellular Ca^{2+} without the opening of voltage-gated Ca^{2+} channels. Subunits $\alpha 7$, $\alpha 8$, and $\alpha 9$ are also the α-bungarotoxin-binding subtypes of neuronal nAChRs.

For the nAChR from muscle, desensitization is minor and probably is not of physiological significance in determining the shape of the synaptic response at the neuromuscular junction. However, for some neuronal nAChRs, desensitization plays a major role in determining the effects of the actions of ACh. Receptors composed of $\alpha 7$, $\alpha 8$, and certain α/β combinations exhibit desensitization time constants of between 100 and 500 ms, whereas others exhibit desensitization constants between 2 and 20 s. Rapid desensitization eliminates that receptor from contributing to synaptic responses until the process is reversed. Given the diverse functions of neuronal nAChRs, the variable rates of desensitization likely play important roles whereby this inherent property of the receptor shapes the physiological response generated from binding ACh.

One Serotonin Receptor Subtype, 5-HT3, is Ionotropic and is a Close Relative of the nAChR

5-HT historically is thought to activate GPCRs (described in more detail later). The 5-HT3 subclass is an exception forming an ionotropic receptor activated by binding 5-HT. 5-HT3 receptors are distributed sparsely on primary sensory nerve endings in the periphery and are also distributed widely in the mammalian CNS. The 5-HT3 receptor is clinically significant because antagonists of 5-HT3 receptors have important applications as antiemetics, anxiolytics, and antipsychotics.

The 5-HT3 receptor is permeable to Na^+ and K^+ ions and is similar in many ways to the nAChR (see evolutionary tree; Fig 8.2) in that both desensitize rapidly and are blocked by tubocurarine. The 5-HT3 receptor is a homomeric complex composed of five copies of the same subunit. Each protein subunit is 487 amino acids long (56 kDa) and has a structure most analogous to the $\alpha 7$ subtype of neuronal nAChRs, which also forms a homo-oligomeric receptor. The 5-HT3 receptor is mostly impermeable to divalent cations. For example, Ca^{2+} is largely excluded from permeation and in fact effectively blocks current flow through the pore, even though the pore size of the channel (7.6 Å) is approximately the same as that for the nAChR (8.4 Å). Apparently, other physical or electrochemical barriers limit the capacity of divalent ions to permeate the 5-HT3 pore. Dose–response studies indicate that at least two ligand-binding sites must be occupied for the channel to open; however, the binding of agonist and/or opening of the channel appears to be approximately 10 times slower than for most other ligand-gated ion channels. The functional significance or physical explanation of this slow opening is not known. The native 5-HT3 receptor also exhibits desensitization (time constant 1–5 s), although the reported desensitization rate varies widely, depending on the methodology used for analysis and the source of receptor. Interestingly, this desensitization can be significantly slowed or enhanced by single amino acid substitutions at a Leu residue in the TM2 segment of the subunit.

$GABA_A$ Receptors Are Related in Structure to nAChRs but Exhibit an Inhibitory Function

Synaptic inhibition in the mammalian brain is mediated principally by GABA receptors. The most widespread ionotropic receptor activated by GABA is designated $GABA_A$. The subunits composing the $GABA_A$ receptor have sequence homology with the nAChR subunit family, and the two families have presumably diverged from a common ancestral gene. The evolutionary relationship of these two receptor families can be appreciated in Fig. 8.2. The $GABA_A$ receptor is composed of multiple subunits, forming a heteropentameric complex

of approximately 275 kDa. Seven different types of subunits are associated with $GABA_A$ receptors and are designated α, β, γ, δ, ε, π and θ. An additional subunit, ρ, is found predominantly in the retina, whereas the other subunits are distributed widely in the brain. Each subunit group also has different subtypes (see Fig 8.2); for example, six different α, three β, three γ, and two other ρ subunits have also been identified. The predicted amino acid sequences indicate that each of these subunits has a molecular mass ranging between 48 and 64 kDa. Like neuronal nAChR, these subunits mix in a heterogeneous fashion to produce a wide array of $GABA_A$ receptors with different pharmacological and electrophysiological properties. The predominant $GABA_A$ receptor in brain and spinal cord is α1, β2, and γ2, with a stoichiometry of two α1s, two β2s, and one γ2. Expression of subunit cDNAs in oocytes indicates that the α subunit is essential for producing a functional channel. The α subunit also appears to contain the high-affinity binding site for GABA.

The ion channel associated with the $GABA_A$ receptor is selective for the anion Cl^-, and the selectivity is provided by amino acid residues strategically positioned in the TM2 near the ends of the pore. When GABA binds to and activates this receptor, Cl^- flows into the cell, producing a hyperpolarization by moving the membrane potential away from the threshold for firing an action potential. The neuronal $GABA_A$ receptor exhibits multiple conductance levels, with the predominant conductance being 27–30 pS.

The $GABA_A$ Receptor Binds Drugs That Affect Its Properties

The $GABA_A$ receptor is an allosteric protein and the binding of a number of compounds modulates its properties. Two well-studied examples are barbiturates and benzodiazepines, both of which bind to the $GABA_A$ receptor and potentiate GABA binding, although they do this by binding to distinct sites. Benzodiazepines produce effects by increasing the frequency of channel opening, while barbiturates prolong channel burst durations. The net result is that in the presence of barbiturates, benzodiazepines, or both, the same concentration of GABA will cause increased inhibition by increasing the amount of Cl^- influx leading to increased hyperpolarization. Benzodiazepine binding is conferred on the receptor by the γ subunit, but the presence of α and β subunits is necessary for the qualitative and quantitative aspects of benzodiazepine binding. The benzodiazepine-binding site appears to lie along the interface between α and γ subunits and only certain subtypes are sensitive to benzodiazepines. Benzodiazepine binding to $GABA_A$ receptors requires α2 or α5 and γ2 or γ3; other subunit combinations are insensitive to benzodiazepines.

Picrotoxin, a potent convulsant compound, binds within the channel pore of the $GABA_A$ receptor and prevents ion flow. Single-channel experiments indicate that picrotoxin either slowly blocks an open channel or prevents the GABA receptor from undergoing a transition into a long-duration open state. Apparently, barbiturates produce similar changes in channel properties, but they potentiate rather than inhibit $GABA_A$ receptor function. Bicuculline, another potent convulsant, appears to inhibit $GABA_A$ receptor channel activity by decreasing the binding of GABA to the receptor. Steroid metabolites of progesterone, corticosterone, and testosterone also have potentiating effects on GABA currents that are similar in many ways to the action of barbiturates; however, the binding sites for steroids and barbiturates are distinct. Finally, penicillin directly inhibits GABA receptor function, apparently by binding within the pore.

The physiological effects of compounds such as picrotoxin, bicuculline, and penicillin are striking. Each of these compounds at a sufficiently high concentration can produce widespread and sustained seizure activity. Conversely, many, but not all, of the sedative properties associated with barbiturates and benzodiazepines can be attributed to their ability to augment inhibition in the brain through enhancing the inhibitory potency of GABA.

Interestingly, ρ-subunit-containing GABA receptors, found in abundance in the retina, are pharmacologically unique. They are resistant to the inhibitory action of bicuculline, although they remain sensitive to blockage by picrotoxin. In addition, these retinal receptors are not sensitive to modulation by barbiturates or benzodiazepines. Thus, ρ-containing receptors are distinct from $GABA_A$ receptors and are similar to receptors earlier designated $GABA_C$.

The Glycine Receptor is Closely Related to the $GABA_A$ Receptor

Glycine receptors are the major inhibitory receptors in the spinal cord and the brain stem. Glycine receptors are similar to $GABA_A$ receptors in that both are ion channels selectively permeable to the anion Cl^-. As one would anticipate, they both lie on the same main branch of the evolutionary tree of ionotropic receptors (Fig. 8.2) Interestingly, only three amino acid replacements in the TM1-TM2 loop and the TM2 segment can change the selectivity of the glycine receptor from anionic to cationic, pointing again to the critical nature of residues in the TM2 segment in determining a receptor's ion selectivity. The overall structure of the glycine receptor is indicative of this similarity in properties. The native complex is approximately 250 kDa and is composed of

two main subunits: α (48 kDa) and β (58 kDa). The receptor appears to be pentameric, most likely composed of three α and two β subunits. The glycine receptor has an open channel conductance of approximately 35–50 pS, similar to that of the $GABA_A$ receptor. The rat poison strychnine is a potent antagonist of the glycine receptor.

Four distinct α subunits and one β subunit of the glycine receptor have been cloned and are highly related (Fig. 8.2). Each exhibits the typical predicted four TMs and is approximately 50% identical to the others at the amino acid level. Expression of a single α subunit in oocytes is sufficient to produce functional glycine receptors, indicating that the α subunit is the pore-forming unit of the native receptor. β subunits play exclusively modulatory roles, affecting, for example, sensitivity to the inhibitory actions of picrotoxin. They are widespread in the brain but not always found in neurons expressing α subunits, suggesting β subunits may serve other functions independent of their association with glycine receptors.

The $GABA_A$ and Glycine Receptor Cluster at Synapses

During maturation of inhibitory synapses, gephyrin is found beneath the postsynaptic membrane and recruits and localizes $GABA_A$ or glycine receptors. Gephyrin appears to serve an analogous function for the stabilization of inhibitory receptors that rapsyn plays for nAChRs (Fig 8.5). Gephyrin is a multidomain protein that interacts with the cytoplasmic domains of the $GABA_A$ or glycine receptor subunits. Like rapsyn, gephyrin interacts with itself and other proteins in addition to the inhibitory receptors and these interactions serve to restrict the lateral mobility of $GABA_A$ and glycine receptors in the plasma membrane. Ultimately these interactions promote the formation of postsynaptic inhibitory specializations.

Certain Purinergic Receptors Are Also Ionotropic

Purinergic chemical transmission is distributed throughout the body and is considered in greater detail in a later section on GPCRs. Purinergic receptors bind to ATP (or other nucleotide analogs) or its breakdown product adenosine (Junger, 2011). ATP is released from certain synaptic terminals in a quantal manner and often is packaged within synaptic vesicles containing another neurotransmitter, the best described being ACh and catecholamines.

One class of ATP-binding purinergic receptors (P2X) has been discovered to be ionotropic, and there are at least seven members of this class (P2X1–P2X7). Note that the P2X receptors reside between the GABA/glycine and the glutamate family of ionotropic receptors on the evolutionary tree (Fig. 8.2). P2X receptors most often mediate a fast depolarizing response in neurons and muscle cells in response to ATP or other nucleotides by the direct opening of relatively nonselective cation channels. cDNAs encoding the P2X receptor indicate that its structure comprises only two transmembrane domains, with some homology in its pore-forming region with K^+ channels and the functional receptors seemed to be composed of three subunits in contrast to five for the nAChR. The P2X7 receptor is also a ligand-gated channel that permits permeation of either anions or cations and even molecules as large as 900 Da in the presence of high (millimolar) nucleotide levels.

Glutamate Receptors Are Derived from a Different Ancestral Gene and Are Structurally Distinct from Other Ionotropic Receptors

Glutamate receptors are widespread in the nervous system, where they are responsible for mediating the vast majority of excitatory synaptic transmission in the brain and spinal cord. Four agonists serve to distinguish different types of glutamate receptors; *N*-methyl-D-aspartate (NMDA), amino-3-hydroxy-5-methylisoxazolepropionic acid (AMPA), kainate, and quisqualate (Hollmann & Heinemann, 1994; Watkins, Krogsgaard-Larsen, & Honore, 1990). A convenient distinction for describing ionotropic glutamate receptors has been to classify them as either NMDA or non-NMDA subtypes, depending on whether they bind the agonist NMDA. Non-NMDA receptors also bind the agonist kainite or AMPA. Both NMDA and non-NMDA receptors are ionotropic. Quisqualate is unique within this group of agonists in having the capacity to activate both ionotropic and GPCR glutamate receptor subtypes (mGluRs; Hollmann & Heinemann, 1994). The nomenclature describing members of the glutamate receptor family has undergone transitions over the years but appears to have converged; receptors binding AMPA are termed GluA1–GluA4; those binding kainate are GluK1–GluK5 and those binding NMDA are GluN1, GluN2A–GluN2D, and GluN3A–GluN3B; and those that activate G-proteins are designated mGluR1–mGluR8 (Traynelis et al., 2010). A family tree highlighting the evolutionary relationship of the glutamate receptors between themselves and other members of the ionotropic receptor family is shown in Figure 8.2.

Non-NMDA Receptors Are a Diverse Family

The initial glutamate receptor clone was termed GluR-K1, and the cDNA encoded a protein with an estimated molecular mass of 99.8 kDa (Hollmann,

O'Shea-Greenfield, Rogers, & Heinemann, 1989). Not long after this original report, several groups (Boulter et al., 1990; Keinanen et al., 1990; Nakanishi, Shneider, & Axel, 1990) independently reported the isolation of families of glutamate receptor subunits, now termed GluA1–GluA4, GluK1–GluK5, and GluD1 and GluD2. Each GluR subunit consists of approximately 900 amino acids and has four predicted membrane-spanning segments (TM1–TM4);

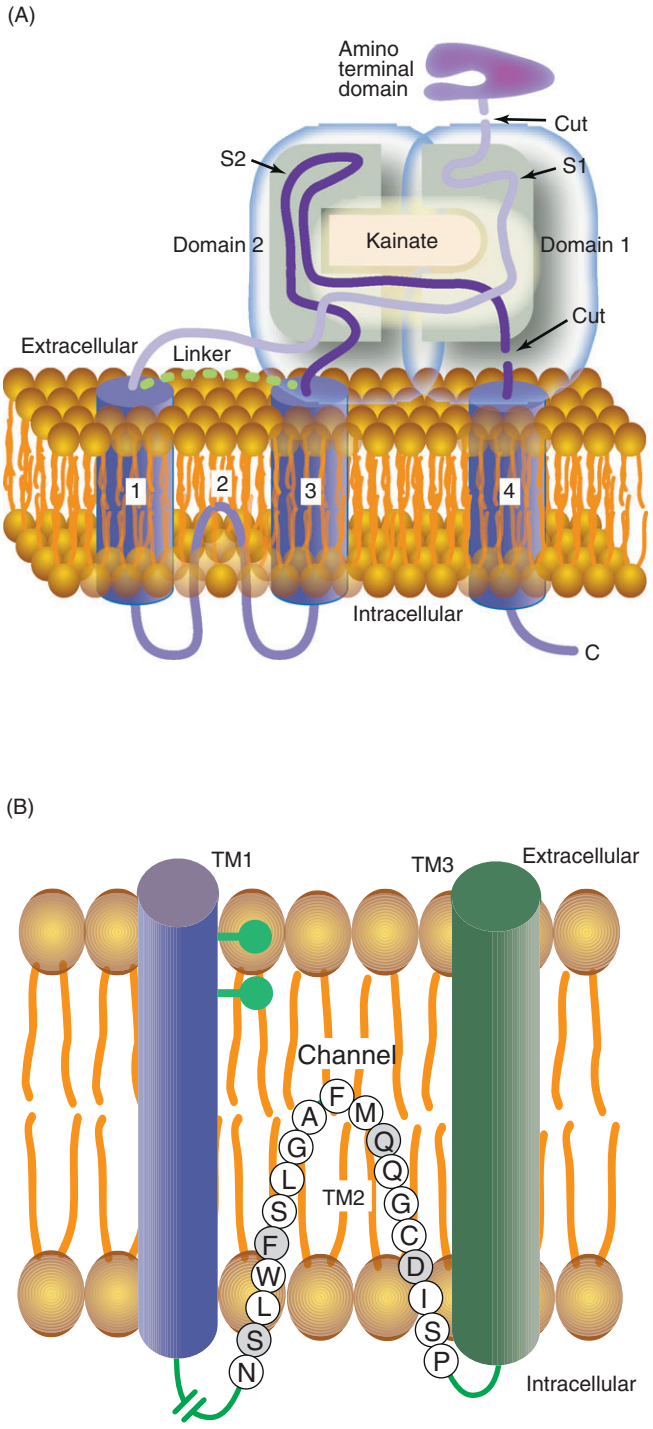


FIGURE 8.6 (A) Model of one of the subunits of the ionotropic glutamate receptor. Ionotropic glutamate receptors have four membrane-associated segments; however, unlike nAChR, only three of them completely traverse the lipid bilayer. TM2 forms a loop and exits back into the cytoplasm. This leaves the large N-terminal region extending into the extracellular space, whereas the C terminus extends into the cytoplasm. Two domains in the extracellular segments associate with each other to form the binding site for transmitter, in this example kainate, a naturally occurring agonist of glutamate. (B) Enlarged area of the predicted structure and amino acid sequence of the TM2 region of the glutamate receptor, GluR3. TM1 and TM3 are drawn as cylinders in the membrane flanking TM2. The residue that determines Ca^{2+} permeability of the non-NMDA receptor is the glutamine residue (Q) highlighted in gray. In NMDA receptors, an asparagine residue at this same position is the proposed site of interaction with Mg^{2+} ions that produce the voltage-dependent channel block. Serine (S) and phenylalanine (F), also shaded in gray, are highly conserved in the non-NMDA receptor family. The aspartate (D) residue is also conserved and is thought to form part of the internal cation-binding site. The break in the loop between TM1 and TM2 indicates a domain that varies in length among ionotropic glutamate receptors. *Adapted from Wo and Oswald (1995).*

however, there is an important distinction in the TM2 domain, making the GluRs distinct from the nAChR family (Figs. 8.6A and 8.6B). Native GluRs are tetrameric, as opposed to the pentameric nAChR family of receptors, and have approximate molecular mass of 600 kDa making them almost twice the size of the nAChR, mostly because of the large extracellular domain where glutamate binds.

Unique Properties of Non-NMDA Receptors Are Determined by Assembly of Different Subunits

When cDNAs encoding these receptors were expressed in either oocytes or HeK-293 cells, application of the non-NMDA receptor agonist AMPA produced substantial inward currents. A striking observation from these expression studies was that when the GluA2 subunit was expressed alone, little stimulated current was obtained, unlike the large currents found when either GluA1 or GluA3 was expressed (Nakanishi et al., 1990; Verdoorn et al., 1991). GluA2 subunits by themselves appear to form poorly conducting receptors. However, when GluA2 is expressed with either GluA1 or GluA3, the behavior of the heteromeric receptor is distinctly different. When GluA1 and GluA3 are expressed alone or together, they produce channels with strong inward rectification. Coexpression of GluA2 with either GluA1 or GluA3 produces a channel with little rectification. Further analyses indicated that GluA1 and GluA3 expressed channels were permeable to Ca^{2+} (Hollmann & Heinemann, 1994). In contrast, any combination of receptor that included the GluA2 subunit produced channels impermeable to Ca^{2+}. The replacement of a single amino acid (Arg for Gln) in TM2 of the GluA2 subunit (see Fig. 8.6B for identification of this amino acid) was shown to control Ca^{2+}-permeability

(Burnashev et al., 1992). Apparently, an Arg at this position blocks Ca^{2+} from traversing the pore formed in the center of the GluR channel.

Functional Diversity in GluRs is Produced by mRNA Splicing and RNA Editing

Analysis of mRNAs encoding GluR subunits indicated that each could be expressed in one of two splice variants, termed flip and flop. These flip and flop modules are small segments (38 amino acids) just preceding the TM4 domain in all four GluA subunits. The receptor channel expressed from these splice variants has distinct properties, depending on which of the two modules is present. Specifically, flop-containing receptors desensitize more during glutamate application. Therefore, GluAs with flop modules express smaller steady-state currents than GluAs with flip modules. Both flip- and flop-containing GluAs are widely expressed in the brain with a few exceptions.

Typically, one finds that there is fidelity in the process of transcribing DNA into mRNA and then into protein—that is, that nucleotides present in the DNA are accurate predictors of the amino acid sequence of the translated protein. However, a novel mechanism in the neuronal nucleus was discovered that edits mRNAs post-transcriptionally, and at least three of the four GluA subunits are subjected to this process (Sommer, Kohler, Sprengel, & Seeburg, 1991). In fact, one of the sites edited is the critical Arg residue regulating Ca^{2+} permeability in the GluA2 subunit. At another edited site, Gly replaces Arg-764 in the GluA2 subunit, and this editing also takes place in GluA3 and GluA4. The Arg-to-Gly conversion at amino acid 764 produces receptors that exhibit significantly faster rates of recovery from the desensitized state. The extent to which other receptors or other protein molecules undergo this form of editing is an area rich for investigation. At a minimum, this editing mechanism produces dramatic differences in the function of GluRs.

Glutamate Receptors Do Not Conform to the Typical Four Transmembrane-Spanning Segment Structure Described for nAChRs

The field of glutamate receptor structure/function is advancing at a rapid pace fostered by the crystallization of the GluA2 receptor (Sobolevsky, Rosconi, & Gouaux, 2009) and report of the 3D structure of the native GluR determined by electron microscopy reconstructions (Nakagawa, Cheng, Ramm, Sheng, & Walz, 2005). These studies reveal the overall architecture of the tetrameric GluR complex and provide some global insights into how glutamate binding leads to channel opening. These models can be additionally refined by the significant amount of work analyzing crystal structures of the isolated GluR ligand-binding domain with and without the presence of bound agonists and antagonists (see Traynelis et al., 2010, for review).

The GluR has a large extracellular domain composed of the amino-terminal domain (ATD) important for receptor assembly and trafficking and the ligand-binding domain (LBD) that serves as the binding site for glutamate (Fig. 8.6A). These domains form a structurally unique dimer arrangement in the extracellular domain that transitions into a fourfold symmetry of the pore-forming region of the TM domains. Unlike the four TM segments of the nAChR that each completely traverses the membrane, the TM2 segment of GluRs forms a reentrant loop within the membrane (Fig. 8.6A) and enters back into the cytoplasm, similar in some ways to the pore-forming domain (P segment) of voltage-activated K^+ channels. An enlargement of this TM2 segment (Fig. 8.6B) highlights the amino acids conserved in the GluRs and further shows the positions of the critical Gln residue responsible for Ca^{2+} permeability of the receptor. Thus, it appears that glutamate receptors are highly divergent from the nAChR receptor family.

Other Non-NMDA GluRs Have Less Well-Characterized Functions

GluK1-GluK5 forms a second non-NMDA receptor subfamily (Fig. 8.2) that is activated by the agonist kainate but whose contribution to producing functionally distinct receptors is less well understood (Traynelis et al., 2010). Their overall structure is similar to that of GluA1–4, and they exhibit about 40% sequence homology; however, their agonist-binding profile and their electrophysiological properties are distinct. GluK1–GluK2 can form either homo- or heterotetramers, but GluK4 and GluK5 require coexpression with one of the GluK1–Gluk3 family members. They are generally expressed at lower levels in the brain than the GluA1–4 subtypes (Hollmann & Heinemann, 1994). GluK4 is expressed at significant concentrations in only two cell types: hippocampal CA3 and dentate granule cells. GluK5 exhibits distinct properties when combined with other GluR subunits. For example, coexpression of GluK2 and GluK5 produces functional receptors that respond to AMPA, although neither subunit itself responds to this agonist.

GluD1 and GluD2 receptors can form homotetrameric receptors, but do not seem to form heteromers with other GluRs. They are also resistant to activation by all known agonists capable of activating other members of the GluR family (Traynelis et al., 2010). Their physiological role remains poorly characterized.

NMDA Receptors Are a Family of Ligand-Gated Ion Channels That Are Also Voltage Dependent

NMDA receptors are involved in aspects of development, learning and memory, and neuronal damage due to brain injury. The particular significance of this receptor to neuronal function comes from two of its unique properties. First, the receptor exhibits associativity. For the channel to be open, the receptor must bind glutamate and the membrane in which it resides must be depolarized. This behavior is due to an Mg^{2+}-dependent block of the receptor at normal membrane resting potentials. Second, the receptor permits a significant influx of Ca^{2+}, and increases in levels of the second messenger Ca^{2+} can activate a variety of processes that alter the properties of the neuron. Excess Ca^{2+} is also toxic, and the hyperactivation of NMDA receptors is thought to contribute to a variety of neurodegenerative disorders.

The NMDA receptor binds a large number of distinct pharmacological compounds. For example, certain hallucinogenic compounds, such as phencyclidine (PCP) and dizocilpine (MK-801), are effective blockers of the ion channel associated with the NMDA receptor (Fig. 8.7). These potent antagonists require the channel to be open to gain access to their binding sites and therefore are referred to as open-channel blockers. They also become trapped when the channel closes and are difficult to wash out once bound. Antagonists for the glutamate-binding site also have been developed, and some of the most well known are AP-5 and AP-7. These and other glutamate-binding site antagonists also produce hallucinogenic effects in humans. NMDA remains a specific agonist for this receptor; however, it is about one order of magnitude less potent than L-glutamate for receptor activation. L-Glutamate is the predominant endogenous neurotransmitter that activates the NMDA receptor; however, L-aspartate can also activate the receptor, as can an endogenous dipeptide in the brain, *N*-acetylaspartylglutamate (Hollmann & Heinemann, 1994).

NMDA Receptor Subunits Show Similarity to Non-NMDA Receptor Subunits

The first cloned NMDA subunit was named NMDAR1 (now termed GluN1), and the deduced amino acid sequence indicated a protein of approximately 97 kDa, similar to other members of the GluR family. Four potential transmembrane domains were identified, and four individual subunits are thought to compose the macromolecular NMDA receptor complex. However, recall that the transmembrane organization of GluR subunits indicates that TM2 forms a reentrant loop in the membrane (Fig. 8.6), and it is assumed that

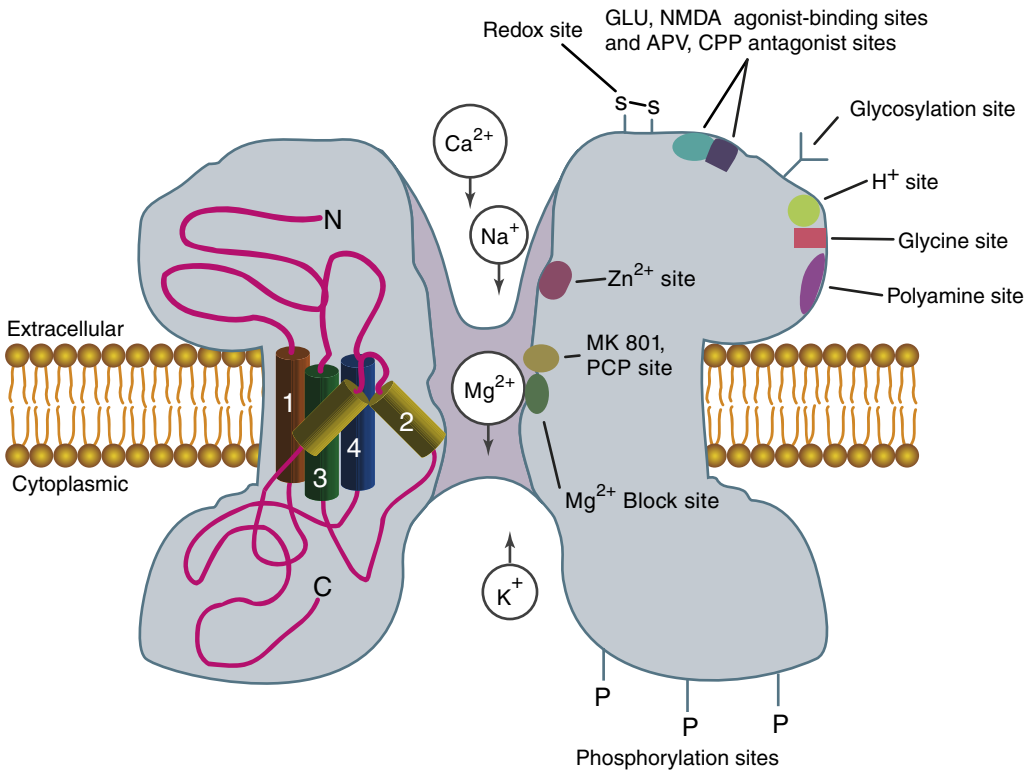


FIGURE 8.7 Diagram of an NMDA receptor highlighting binding sites for numerous agonists, antagonists, and other regulatory molecules. The location of these sites is a crude approximation for the purpose of discussion. *Adapted from Hollmann and Heinemann (1994).*

NMDA receptor subunits will also follow this structural topology (Fig. 8.7). The TM2 segment of each subunit clearly lines the pore of the NMDA receptor channel. In fact, a single Asn residue in TM2, analogous to that in the GluA2 subunit, regulates the Ca^{2+} permeability of the NMDA receptor.

Three of the best-characterized facets of the NMDA receptor were found when the GluN1 subunit was initially expressed in oocytes, although currents were unusually small. These characteristics are (1) a Mg^{2+}-dependent, voltage-sensitive ion channel block, (2) a glycine requirement for effective channel opening, and (3) Ca^{2+} permeability. However, these initial expression studies were complicated by the apparent endogenous expression of other GluN-like subunits. No glutamate-stimulated current is found when GluN1 is expressed in mammalian cells and the NMDAR is now recognized to exist in the nervous system as obligate heteromers of GluN1 mixed with other subunits (Traynelis et al., 2010).

Several other members of the NMDA receptor family have been cloned (GluN2A–2D and GluN3A and GluN3B), and their deduced primary structures are highly related to each other and to other GluRs (see Fig. 8.2). These GluN receptor subunits do not form channels when expressed singly or in combination unless they are coexpressed with GluN1. Apparently, GluN1 serves an essential role for the formation of a functional pore by which activation of NMDA receptors permits the flow of ions. The native NMDA receptor is thought to be composed of two GluN1 subunits and either two GluN2 subunits or a combination of GluN2 and GluN3 subunits. The GluN1 or GluN3 subunits provide the glycine-binding site for the receptor, while the GluN2 subunits contain the glutamate binding sites (Traynelis et al., 2010). The C-terminal domains of GluN2A–GluN2D are quite large relative to the GluN1 C-terminus and appear to play roles in altering channel properties and in determining the subcellular localization of the receptors. All the NMDAR subunits have an Asn residue at the critical point in the TM2 domain essential for producing Ca^{2+} permeability. This Asn residue also appears to form at least part of the binding site for Mg^{2+}, indicating that the sites for Mg^{2+} binding and Ca^{2+} permeation overlap.

The distribution of GluN2 subunits generally is more restricted than the homogeneous distribution of GluN1, with the exception of GluN2A, which is expressed throughout the nervous system. GluN2C is restricted mostly to cerebellar granule cells, whereas 2B and 2D exhibit broader distributions. As noted, the large size of the C terminus of the GluN2 subunit suggests a potential role in association with other proteins, possibly to target or restrict specific NMDA receptor types to areas of the neuron (see below). Mechanisms related to receptor targeting now are becoming understood and will clearly play major roles in determining the efficacy of synaptic transmission.

Additional NMDA Receptor Diversity Occurs through RNA Splicing

At least eight splice variants now have been identified for the GluRN1 subunit and these variants produce differences, ranging from subtle to significant, in the properties of the expressed receptor (Hollmann & Heinemann, 1994; Traynelis et al., 2010). For example, GluN1 receptors lacking a particular N-terminal insert due to alternative splicing exhibit enhanced blockade by protons and exhibit responses that are potentiated by Zn^{2+} in micromolar concentrations. Zn^{2+} classically has been described as an NMDA receptor antagonist that significantly blocks its activation. Clearly, the particular splice variant incorporated into the receptor complex affects the types of physiological response generated. Spermine, a polyamine found in neurons and in the extracellular space, also slightly increases the amplitude of NMDA responses, and this modulatory effect also appears to be associated with a particular splice variant. The physiologic role of spermine in regulating NMDA receptors remains unclear.

NMDA Receptors Exhibit Complex Channel Properties

The biophysical properties of the NMDA receptor are complex. Single-channel conductance has a main level of 50 pS; however, subconductances are evident, and different subunit combinations and splice variants produce channels with distinct single-channel properties. A binding site for the Ca^{2+}-binding protein calmodulin also has been identified on the GluN1 subunit. Binding of Ca^{2+}–calmodulin to NMDA receptors produces a fourfold decrease in open-channel probability. Ca^{2+} influx through the NMDA receptor could induce calmodulin binding and lead to an immediate short-term feedback inhibition, decreasing ion flow through the receptor.

Glutamate Receptors Cluster at Synapses

Glutamate receptors are concentrated at excitatory synapses through interactions with underlying scaffolding molecules (Fig. 8.8). These scaffolding molecules and a host of other structural and signalling proteins form an electron dense structure at excitatory synapses called the postsynaptic density (PSD). One of the best known of these scaffolds is PSD95, a multidomain protein enriched at excitatory synapses that helps organize receptors and signalling molecules. Through one of its multiple PDZ domains, PSD95 binds directly to the

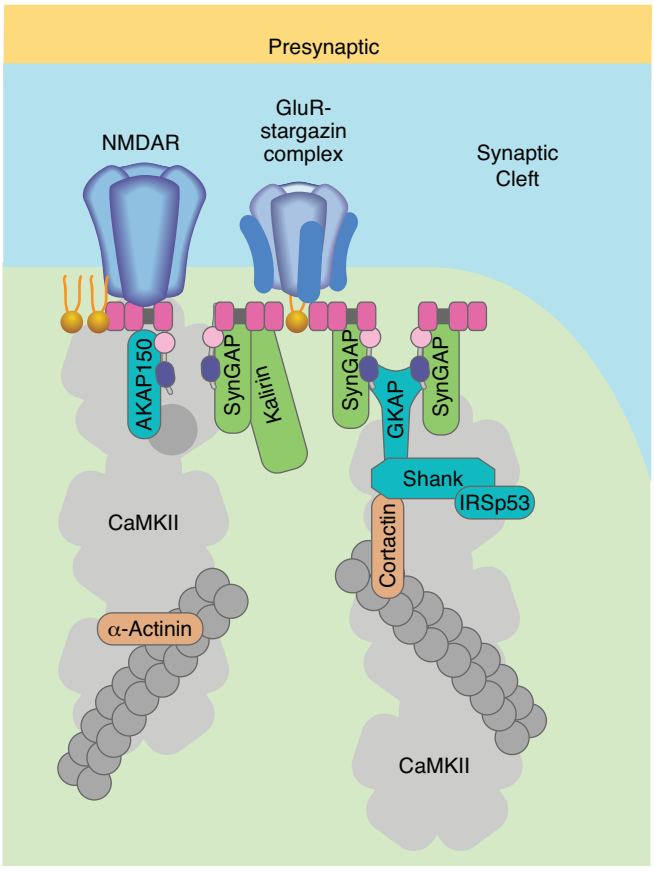


FIGURE 8.8 Diagram of glutamate receptor clustering at an excitatory synapse. The NMDA receptor interacts directly with PSD-95 through binding to one of PSD-95's three PDZ domains (the PDZ domains of PSD-95 are shown as pink squares). The AMPAR is associated with a protein called stargazin and stargazin interacts with one of the PDZ domains of PSD-95. Only a few of the many other signaling and scaffolding proteins at excitatory synapses are shown. AKAP150 is A-kinase anchoring protein of 150 kDa that binds to protein kinase A and other proteins, SynGAP is an abundant synaptic associated Ras GTP-ase activating protein that interacts with PSD-95, GKAP is a guanylate kinase associated protein that interacts with PSD-95, CaMKII is an abundant Ca^{2+}/calmodulin-activated protein kinase that interacts directly with the NMDAR. CaMKII also interacts with itself and with α-actinin, which is an actin-binding protein. This web of protein-protein interactions forms the electron dense structures called the postsynaptic densities visible in electon micrographs of excitatory synapses. *Adapted from Sheng and Hoogenraad (2007).*

cytoplasmic domain of GluN2 subunits and anchors them tightly to the PSD (Sheng & Hoogenraad, 2007). In fact, biochemically isolated PSDs stripped of the overlying membrane retain GluNs indicating the interaction is mediated by tight binding. A number of other PSD enriched proteins have been identified to bind to the C-terminal domain of GluN receptor subunits, including the Ca^{2+}/calmodulin-activated protein kinase CaMKII (Fig. 8.8).

GluA receptors do not interact directly with PSD-95, and their relative concentration in the PSD is dynamic. The up- and down-regulation of GluA receptor number in the postsynaptic membrane alters the efficiency of synaptic transmission and provides a mechanism for the expression of plasticity at excitatory synapses. One accessory protein factor important for receptor trafficking is the protein stargazin that was recently discovered as a structural component of isolated GluA receptors. Stargazin is one member of a family of proteins called TARPs for transmembrane AMPA receptor regulatory proteins that are important for the maturation and targeting of GluA receptors to excitatory synapses (Fig. 8.8). TARPs bind tightly to GluA receptors and they can be identified as additional protein density in electron microscopic reconstructions of isolated GluA receptors (Nakagawa et al., 2005). This situation is analogous in most respects to that described for rapsyn binding to the nAChR (Fig. 8.5). However, TARPs are themselves membrane-spanning proteins, having four predicted TM domains. TARPs also have a domain that interacts with PSD-95 (or other PDZ containing proteins), and it is this interaction that leads to the association of GluA receptors with PSDs. TARPs also affect the maturation and channel properties of GluA receptors in addition to their role in synaptic targeting. Interestingly, TARP binding decreases when glutamate binds to the receptor suggesting interplay exists between receptor use and TARP mediated localization. There are other well-documented proteins that interact with the C-terminal domain of GluA receptor subunits, some in a subtype specific manner. This list includes GRIP/ABP, PICK-1, NSF, and SAP-97, and these proteins have likewise been implicated in the maturation and trafficking of GluA receptors. It is also well documented that specific residues in the C-terminal domain of GluA receptors are phosphorylated and phosphorylation regulates channel properties and surface expression (Traynelis et al., 2010). These post-translational modifications and interactions with other proteins must work in concert for the proper localization and stability of non-NMDA receptors at synapses.

Summary

A general model for ionotropic receptors has emerged mainly from analyses of the nAChR. Ionotropic receptors are large membrane-bound complexes generally composed of five subunits. The subunits each have four transmembrane domains, and the amino acids in TM2 form the lining of the pore. Transmitter binding induces rapid conformational changes that are translated into an increase in the diameter of the pore, permitting ion influx. Cation or anion selectivity is obtained through the coordination of specific negatively or positively charged amino acids at strategic locations in the receptor pore. Glutamate receptors are structurally distinct from

the nAChR family of receptors and a major advancement in defining these differences came from the successful crystallization of the GluR2 receptor. This and previous results showed that the TM2 domain of glutamate receptors forms a hairpin instead of traversing the membrane completely, causing the remainder of the receptor to adopt a different subunit topology than that described for the nAChR family. Glutamate receptors are also composed of four, not five, subunits. These differences are perhaps not surprising given that the nAChR family and the glutamate receptor family have arisen from two different ancestral genes.

G-PROTEIN COUPLED RECEPTORS

The number of members in the G-protein coupled receptor family is enormous, with over 800 identified from sequencing of the human genome. Historically, the term *metabotropic* was used to describe the fact that intracellular metabolites are produced when these receptors bind ligand. However, there are now clearly documented cases where the activation of "metabotropic" receptors does not produce alterations in metabolites but instead produces their effects by interacting with G-proteins that directly alter the behavior of ion channels. Thus, these receptors are now referred to as G-protein-coupled receptors (GPCRs). Currently, the GPCR family can be divided into three subfamilies on the basis of their structures: (1) the rhodopsin-adrenergic receptor subfamily (class A), (2) the secretin-vasoactive intestinal peptide receptor subfamily (class B), and (3) the metabotropic glutamate receptor subfamily (class C).

When a GPCR is activated, it couples to a G-protein initiating the exchange of GDP for GTP, activating the G-protein (Fig. 8.1B). Activated G-proteins then couple to downstream effectors to alter the activity of other intracellular enzymes or ion channels. Many of the G-protein target enzymes produce diffusible second messengers that stimulate further downstream biochemical processes, including the activation of protein kinases (see Chapter 9). Time is required for each of these coupling events, and the effects of GPCR activation are typically slower in onset than those observed following the activation of ionotropic receptors. Because there is a lifetime associated with each intermediate, the effects produced by GPCR activation are also longer in duration than those produced by the activation of ionotropic receptors. Most small neurotransmitters, such as ACh, glutamate, 5-HT, and GABA, can bind to and activate both ionotropic and GPCRs. Thus, each of these transmitters can induce both fast responses (milliseconds), such as typical excitatory or inhibitory postsynaptic potentials, and slow-onset and longer duration responses (from tenths of seconds to, potentially, hours). These effects span a broad range of time domains providing the nervous system with a rich source for temporal information processing that is subject to constant modification.

GPCR Structure Conforms to a General Model

A GPCR typically consists of a single polypeptide with a generally conserved structure; however, there are now clear examples where functional GPCRs are composed of dimers. The receptor subunits contain seven helical segments that wrap back and forth through the membrane (Fig. 8.9). GPCRs are homologous to rhodopsin, and detailed structural information on rhodopsin was originally used to provide a framework for describing other GPCR structures. However, there are now crystal structures available for the human adenosine (A2a) receptor, the human D3 receptor, the human β2 adrenoreceptor, and the avian β1 adrenoreceptor that provide important comparisons to describe the unique and conserved features of GPCR structure.

The most conserved feature of GPCRs is the seven membrane-spanning segments; however, other generalities can be made about their structure. The N terminus of the receptor extends into the extracellular space, whereas the C terminus resides within the cytoplasm (Fig. 8.9A). Each of the seven TM domains between N and C termini consists of approximately 24 mostly hydrophobic amino acids. These seven domains associate together to form an oblong ring within the plasma membrane (Fig. 8.9B). Between each transmembrane domain is a loop of amino acids of various sizes. The loops connecting TM1 and TM2, TM3 and TM4, and TM5 and TM6 are intracellular and are labeled i1, i2, and i3, respectively, whereas those between TM2 and TM3, TM4 and TM5, and TM6 and TM7 are extracellular and are labeled e1, e2, and e3, respectively (see Fig. 8.9A).

The Neurotransmitter-Binding Site is Buried in the Core of the Receptor

The neurotransmitter-binding site for class A GPCRs resides within a pocket formed in the center of the seven membrane-spanning segments (Fig. 8.9). In the βAR, this pocket resides approximately 11 Å into the hydrophobic core of the receptor, placing the ligand-binding site within the plasma membrane lipid bilayer (Kobilka, 1992; Rosenbaum, Rasmuseen & Kobilka, 2009). Strategically positioned charged and polar residues in the membrane-spanning segments point inward into a central pocket that forms the binding site for the ligand. For example, Asn residues in the second and third segments, two Ser residues in the fifth segment, and a Phe residue in the sixth segment provide major contact

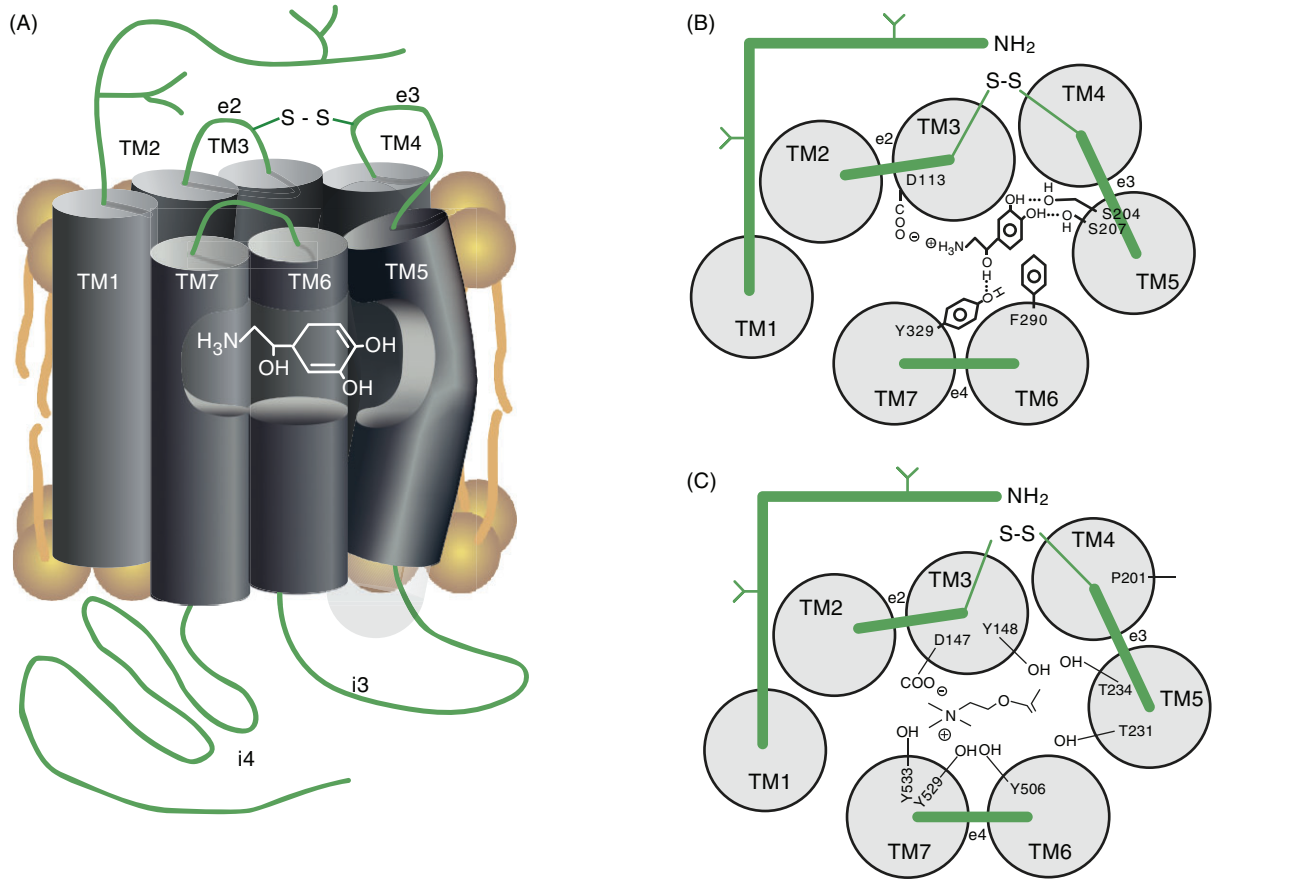


FIGURE 8.9 (A) Diagram showing the approximate position of the catecholamine-binding site in the βAR. The transmitter-binding site is formed by amino acids whose side chains extend into the center of the ring produced by the seven transmembrane domains (TM1–TM7). Note that the binding site exists at a position that places it within the plane of the lipid bilayer. (B) A view looking down on a model of the βAR identifying residues important for ligand binding. The seven transmembrane domains are represented as gray circles labeled TM1 though TM7. Amino acids composing the extracellular domains are represented as green bars labeled e1 through e4. The disulfide bond (–S–S–) that links e2 to e3 is also shown. Each of the specific residues indicated makes stabilizing contact with the transmitter. (C) A view looking down on a model of the mAChR identifying residues important for ligand binding. Stabilizing contacts, mainly through hydroxyl groups (-OH), are made with the transmitter on four of the seven transmembrane domains. The chemical nature of the transmitter (i.e., epinephrine versus Ach) determines the type of amino acids necessary to produce stable interactions in the receptor-binding site (compare B and C). *Adapted from Strosberg (1990).*

points in the βAR-binding site for the transmitter (Fig. 8.9B; Kobilka, 1992). Replacing the Asp in TM3 with a Glu reduced transmitter binding by more than 100-fold, and replacement with a less conserved amino acid, such as Ser, reduces binding by more than 10,000-fold. Two Ser residues in TM5 are also essential for efficient transmitter binding and receptor activation, as is an Asp residue in TM2 and a Phe residue in TM6. In total, the two Asp, the two Ser, and the Phe residues are highly conserved in all receptors that bind catecholamines. Variations in the amino acids at these five positions appear to provide specificity between the binding of different transmitters to the individual GPCRs.

The neurotransmitter-binding site of mAChRs, like that of the β2AR, has been investigated in great detail (Fig. 8.9C). The Asp residue in TM3 is also critical for ACh binding to mAChRs. Mutagenesis studies indicate important roles for Tyr and Thr residues in TM3, TM5, TM6, and TM7 in contributing to the ligand-binding site for ACh. Interestingly, many of these mutations do not affect antagonist binding, indicating that distinct sets of amino acids participate in binding agonists and antagonists. When the transmembrane domains are examined from a side view (Fig. 8.9A), the key amino acids implicated in agonist binding lie at about the same level within the core of the receptor structure, buried approximately 10–15 Å from the surface of the plasma membrane. An additional amino acid identified as essential for agonist binding of the mAChR is a Pro residue in TM4 (P201 in Fig. 8.9C). This residue is also highly conserved among GPCRs, and structural predictions suggest that it affects ligand binding not by interacting with agonist directly but by stabilizing a conformation essential for high-affinity binding. Structural predictions

also place this Pro residue in the same plane as the Asp, Tyr, and Thr residues that form the ligand-binding site of the mAChR.

Transmitter Binding Causes a Conformational Change in the Receptor and Activation of G-Proteins

Proposed models for GPCR activation assume that the receptor can isomerize spontaneously between inactive and active states although there is significant evidence of additional intermediate states along this pathway. Agonist binding stabilizes the active conformation and shifts the equilibrium toward the active form, and G-protein activation ensues. In the absence of agonist, the inactive state of GPCRs is favored, and little G-protein activation occurs. This kinetic model indicates that agonist binding is not necessary for the receptor to undergo transition into the active state; instead, it stabilizes the activated state of the receptor. This model is supported by observations of both spontaneously arising and engineered mutants of the βAR and αAR that can express constitutive activity in the absence of agonist. The mutants stabilize the active conformation of the molecule in a state more similar to the agonist-bound form of the receptor, leading to productive interactions with G proteins. Stabilization of intermediate states by pharmacological agents can explain how certain drugs produce full agonist, partial agonist, neutral antagonist or inverse agonist behavior (Rosenbaum et al., 2009).

The Third Intracellular Loop Forms a Major Determinant for G-Protein Coupling

Extensive studies using site-directed mutagenesis and the production of chimeric molecules have revealed the domains and amino acids essential for G-protein coupling to GPCRs. Receptor domains within the second (i2) and third (i3) intracellular loops (Fig. 8.9) appear largely responsible for determining the specificity and efficiency of coupling for adrenergic and muscarinic cholinergic receptors and are the likely sites for G-protein coupling of the entire GPCR family. In particular, the 12 amino acids of the N-terminal region of the third intracellular loop significantly affect the specificity of G-protein coupling. Other regions in the C terminus of the third intracellular loop and the N-terminal region of the C-terminal tail appear to be more important for determining the efficiency of G-protein coupling than for determining its specificity (Kobilka, 1992). The third intracellular loop varies enormously in size among the different G-protein-coupled receptors, ranging from 29 amino acids in the substance P receptor to 242 amino acids in the mAChR (Strader, Fong, Tota, Underwood, & Dixon, 1994). The intracellular loop connecting TM5 and TM6 is the main point of receptor coupling to G-proteins, and ligand binding to amino acids in TM5 and TM6 may be responsible for triggering the G-protein–receptor interaction by transmitting a conformational change to the third intracellular loop.

Specific Amino Acids Are Involved in Transducing Transmitter Binding into G-Protein Coupling

Residues associated with transmitting the conformational change induced by ligand binding to the activation of G-proteins have been investigated with the use of mAChRs. These studies revealed that an Asp residue in TM2 is important for the receptor activation of G-proteins, and altering the Asp by site-directed mutagenesis has a major negative effect on G-protein–receptor activation. A Thr residue in TM5 (T231 in Fig. 8.9C) and a Tyr residue in TM6 (Y506 in Fig. 8.9C) are also essential. Because these residues are connected by i3, they are assumed to play fundamental roles in transmitting the conformational change induced by ligand binding to the area of i3 essential for G-protein coupling and activation. When mutated, a Pro residue on TM7 (part of a conserved AsnProXXTyr motif) produces a major impairment in the ability of the TM3 segment to induce G-protein coupling. The specifics of these earlier mutagenesis studies have largely been confirmed by comparison of different crystallized forms of the GPCRs and the term "rotamer toggle switch" is used to capture the conformational transition from the inactive to active conformations. There is also a highly conserved sequence AspGluArgTyr in TM3 and a Glu reside in TM6 that forms a series of interactions termed the "ionic lock" that help stabilize the inactive conformation (Rosenbaum et al., 2009). Disruption of this interaction is thought to be an important step in inducing the active conformation capable of productive G-protein interactions.

As mentioned earlier, GPCRs are often single polypeptides; however, they are clearly separable into distinct functional domains. For example, β2AR can be physically split, with the use of molecular techniques, into two fragments: one fragment containing TM1–TM5 and the other containing TM6 and TM7. In isolation, neither of these fragments can produce a functional receptor; however, when coexpressed in the same cell, functional β2ARs that can bind ligand and activate G-proteins are produced. This remarkable experiment indicates that physical connection in the primary sequence is not essential for producing functional β2ARs and also emphasizes the contribution of domains in the separate fragments (TM1–TM5 and TM6 and TM7) to both ligand binding and G-protein coupling.

GPCRs Also Exist as Homo- or Hetero-oligomers

The observation that GPCRs can be physically split through genetic engineering, but produced functional channels when recombined, provided the first hint that full-length GPCRs might also oligomerize with each other into functional molecules. A test of this hypothesis was accomplished by making chimeric receptors composed of TM1–5 of the α2-AR and TM6 and 7 of the m3 muscarinic receptors and vice versa. When either of these chimeric molecules was expressed in isolation, neither formed a functional receptor. However, when coexpressed, receptors were formed that bind both muscarinic and adrenergic ligands and ligand binding led to functional activation of downstream effectors. Through domain swapping, the ligand-binding sites for both receptor ligands were reconstituted by oligomerization of the two chimeric receptors into one bifunctional chimeric dimer. It is now clear that oligomerization of GPCRs into homo- or hetero dimers is adding a new layer of complexity and diversity to the study of these receptors. Important functional consequences could relate to alterations in (1) ligand binding, (2) efficiency and specificity of coupling to downstream effectors, (3) subcellular localization, and (4) receptor desensitization. The evolving and apparently widespread nature of direct receptor–receptor interactions leads one to believe that our current understanding of GPCRs and their biological impact will be undergoing continual modifications for many years to come.

G-Protein Coupling Increases Affinity of the Receptor for Neurotransmitter

The affinity of a GPCR for agonist increases when the receptor is coupled to the G-protein. This positive feedback effectively increases the lifetime of the agonist-bound form of the receptor by decreasing the dissociation rate of the agonist. An excellent demonstration of this effect comes from studies using engineered βAR receptors that are constitutively active in their ability to couple to G-proteins. These mutant receptors show a significantly increased affinity for agonists. When the G-protein dissociates, the agonist-binding affinity of the receptor returns to its original state. Changes induced by ligand binding apparently stabilize the receptor in a conformation with both higher affinity for ligand and higher affinity for coupling to G-proteins.

Specificity and Potency of G-Protein Activation Are Determined by Several Factors

GPCRs associate with G-proteins to transduce ligand binding into intracellular effects. This coupling step can lead to diverse responses, depending on the type of G-protein and the type of effector enzyme present. Ligand binding to a single subtype of GPCR can activate multiple G-protein-coupled pathways. For example, activated α2ARs have been shown to couple to as many as four different G-proteins in the same cell (Strader et al., 1994). Some of the specificity for G-protein activation can be determined by the specific conformations assumed by the receptor, and a single receptor can assume multiple conformations. For example, α2ARs can isomerize into at least two states. One state interacts with a G-protein that couples to phospholipase C, and a second state interacts with G-proteins that couple to both phospholipase C and phospholipase A_2. Thus, a single GPCR can produce a diversity of responses, making it difficult to assign specific biological effects to individual receptor subtypes in all settings.

Activated GPCRs are free to couple to many G-protein molecules, permitting a significant amplification of the initial transmitter-binding event. This catalytic mechanism is referred to as "collision coupling," whereby a transient association between the activated receptor and the G-protein is sufficient to produce the exchange of GDP for GTP, activating the G-protein. For example, because adenylyl cyclase appears to be tightly coupled to the G-protein, the rate-limiting step in the production of cAMP is the number of successful collisions between the receptor and the G-protein. A constant GTPase activity hydrolyzes GTP, bringing the G-protein and therefore the adenylyl cyclase back to the basal state. The apparent maximal rate is achieved when all the G-protein–cyclase complexes have become activated or, more accurately, when the rate of formation is maximal with respect to the rate of GTP hydrolysis. Additionally, the important process of receptor desensitization can also regulate the number of receptors capable of productive G-protein interactions.

Receptor Desensitization is a Built-in Mechanism for Decreasing the Cellular Response to Transmitter

Desensitization is an important process whereby cells can decrease their sensitivity to a particular stimulus to prevent saturation of the system (Kobilka, 1992; Rosenbaum et al., 2009). For GPCRs, desensitization is defined as an increase in the concentration of agonist required to produce half-maximal stimulation of, for example, adenylyl cyclase. In practical terms, desensitization of receptors produces less response for a constant amount of transmitter.

There are two known mechanisms for desensitization. One mechanism is a decrease in response brought

about by the covalent modifications produced by receptor phosphorylation and is quite rapid (seconds to minutes). The other mechanism is the physical removal of receptors from the plasma membrane (through a mechanism of receptor-mediated endocytosis) and tends to require greater periods of time (minutes to hours). The latter process can be either reversible, where receptors are sequestered but are recycled to the plasma membrane or irreversible, where the receptors are removed from the cell through degradation (Fig. 8.10).

The Rapid Phase of GPCR Desensitization is Mediated by Receptor Phosphorylation

Desensitization of the βAR appears to involve at least three protein kinases: PKA, PKC, and a G-protein receptor kinase (GRK). Phosphorylation of ARs by PKA does not require that the agonist be bound to the receptor and appears to be a general mechanism by which the cell can reduce the effectiveness of all receptors, independent of whether they are in the agonist-bound or unbound state. This process is also referred to as heterologous desensitization because the receptor does not require bound-agonist. PKA and PKC phosphorylate sites on the third intracellular loop and possibly the C-terminal cytoplasmic domain. Phosphorylation of these sites interferes with the ability of the receptor to couple to G-proteins, thus producing the desensitization.

GRKs can also phosphorylate GPCRs and lead to receptor desensitization. Six members of the GRK family of kinases have been identified: rhodopsin kinase (GRK1), βARK (GRK2), and GRK3 through GRK6 (Huang & Tesmer, 2011; Premont, Inglese, & Lefkowitz, 1995). GRK2 is a Ser and Thr specific protein kinase initially identified by its capacity to phosphorylate the βAR. GRK2 phosphorylates only the agonist-bound form of the receptor, usually when agonist concentrations reach the micromolar level, as typically found in the synaptic cleft. This process is referred to as homologous desensitization because the regulation is specific for those receptor molecules that are in the transmitter-bound state. Phosphorylation of βAR by GRK2 does not interfere with coupling to G-proteins directly. Instead, an additional protein, arrestin, binds the GRK2-phosphorylated form of the receptor, thus blocking receptor–G-protein coupling (Fig. 8.10). This process is analogous to the desensitization of the light-sensitive receptor molecule rhodopsin produced by GRK1 phosphorylation and the binding of arrestin. Phosphorylation sites on βAR for GRK2 reside on the C-terminal cytoplasmic domain and are distinct from those phosphorylated by PKA.

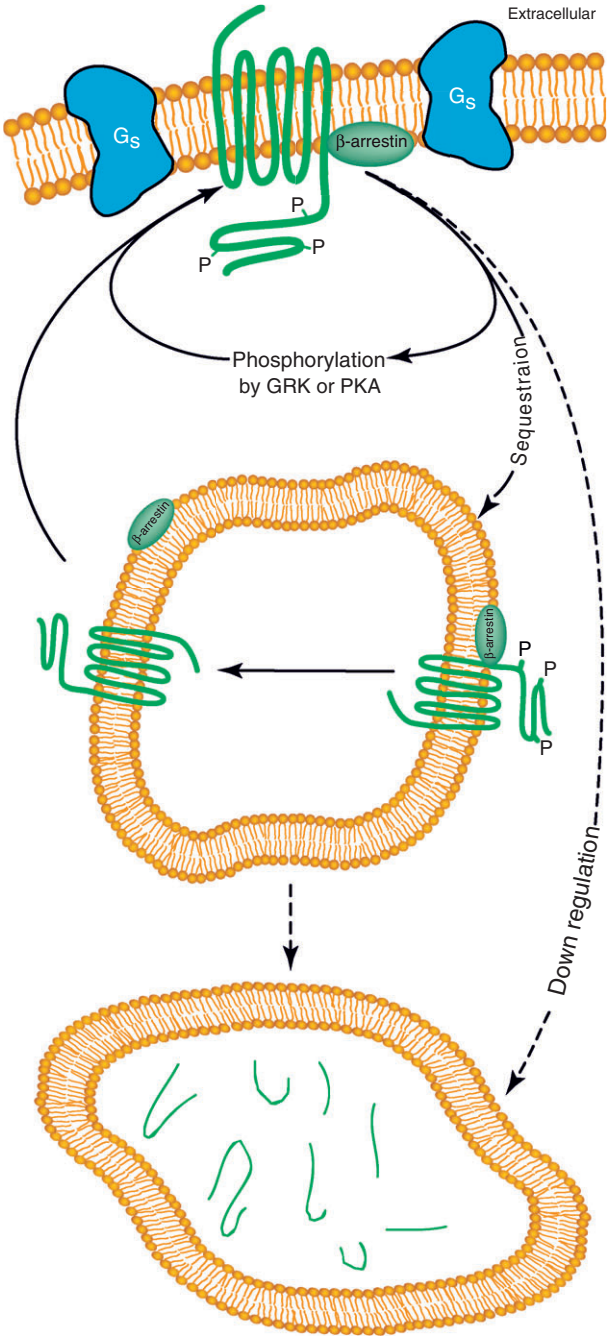


FIGURE 8.10 Intracellular pathways associated with desensitization of GPCRs. GPCRs are phosphorylated (noted with P) on their intracellular domains by PKA, GRK, and other protein kinases. The phosphorylated form of the receptor can be removed from the cell surface by a process called sequestration with the help of the adapter protein β-arrestin; thus fewer binding sites remain on the cell surface for transmitter interactions. In intracellular compartments, the receptor can be dephosphorylated and returned to the plasma membrane in its basal state. Alternatively, phosphorylated receptors can be degraded (down regulated) by targeting to a lysosomal organelle. Degradation requires replenishment of the receptor pool through new protein synthesis. *Adapted from Kobilka (1992).*

The cycle of homologous desensitization starts with the activation of a GPCR, which induces activation of G-proteins and dissociation of the $\beta\gamma$ subunit complex from α subunits. At least one role for the $\beta\gamma$ complex appears to be to bind to GRKs, which leads to their recruitment to the membrane in the area of the locally activated G-protein–receptor complex. The recruited GRK is then activated, leading to phosphorylation of the agonist-bound receptor and subsequent binding of arrestin. Arrestin binds to the same domains on the receptor necessary for coupling to G-proteins, thus terminating the actions of the activated receptor (Fig. 8.10). The ensuing process of sequestration follows GPCR phosphorylation and arrestin binding.

Desensitization Can Also Be Produced by Loss of Receptors from the Cell Surface

Desensitization of GPCRs is also produced by removal of the receptor from the cell surface. This process can be either reversible (sequestration, or internalization) or irreversible (down regulation). Sequestration is the term used to describe the rapid (within minutes) but reversible endocytosis of receptors from the cell surface after agonist application (Fig. 8.10). Neither G-protein coupling nor receptor phosphorylation appears to be essential for this process, but phosphorylation by GRKs clearly enhances the rate of sequestration. The binding of arrestin to the phosphorylated receptor also enhances sequestration (Ferguson et al., 1996; Huang & Tesmer, 2011). Thus, arrestin binding appears to promote not only rapid desensitization by disrupting the receptor–G-protein interaction but also receptor sequestration. Receptor cycling through intracellular organelles is an important trafficking mechanism that leads to an enhanced rate of dephosphorylation of the phosphorylated receptor, returning it to the cell surface in its basal state (Fig. 8.10).

Down regulation occurs more slowly than sequestration and is irreversible (Fig. 8.10). The early phase (within 4h) may involve both a PKA-dependent and a PKA-independent process. This early phase of down regulation is apparently due to receptor degradation after endocytotic removal from the plasma membrane. The later phases (>14h) of down regulation appear to be further mediated by a reduction in receptor biosynthesis through a decrease in the stability of the receptor mRNA and a decreased transcription rate.

Other Post-translational Modifications Are Required for Efficient GPCR Function

Like many proteins expressed on the cell surface, GPCRs are glycosylated, and the N-terminal extracellular domain is the site of carbohydrate attachment. Glycosylation does not appear to be essential to the production of a functional ligand-binding pocket (Strader et al., 1994), although prevention of glycosylation may decrease membrane insertion and alter intracellular trafficking of the β2AR.

Another important structural feature of most GPCRs is the disulfide bond formed between two Cys residues present on the extracellular loops (e2 and e3; Fig. 8.9). This disulfide bond stabilizes a restricted conformation of the mature receptor by covalently linking the two extracellular domains, and this conformation favors ligand binding. Disruption of this disulfide bond significantly decreases agonist binding (Kobilka, 1992). A third Cys residue, in the C-terminal domain of GPCRs (in i4; Fig. 8.9A), serves as a point for covalent attachment of a fatty acid (often palmitate). Fatty acid attachment stabilizes interactions between the C-terminal domain of a GPCR and the membrane.

GPCRs All Exhibit Similar Structures

The family of GPCRs exhibits structural similarities that permit the construction of "trees" describing the degree of their evolutionary kinship (Fig. 8.11). Some remarkable relations become evident in such an analysis. For example, the D1 and D5 subtypes of DA receptors are related more closely to the AR than to the D2, D3, and D4 DA receptors. The similarities and differences among GPCR families are highlighted in the remainder of this chapter and their evolutionary relatedness to each other can be seen in Figure 8.11.

Muscarinic ACh Receptors

Muscarine is a naturally occurring plant alkaloid that binds to muscarinic subtypes of the AChRs to activate them. mAChRs play a dominant role in mediating the actions of ACh in the brain, indirectly producing both excitation and inhibition through binding to a family of unique GPCRs. mAChRs are found both presynaptically and postsynaptically, and, ultimately, their main neuronal effects appear to be mediated through alterations in the properties of ion channels. Presynaptic mAChRs take part in important feedback loops that regulate ACh release. ACh released from the presynaptic terminal can bind to mAChRs on the same nerve ending, thus activating enzymatic processes that modulate subsequent neurotransmitter release. This modulation is typically an inhibition; however, activation of m5AChRs produces an enhancement in subsequent release. These autoreceptors are an important regulatory mechanism for the short-term (milliseconds to seconds) modulation of neurotransmitter release.

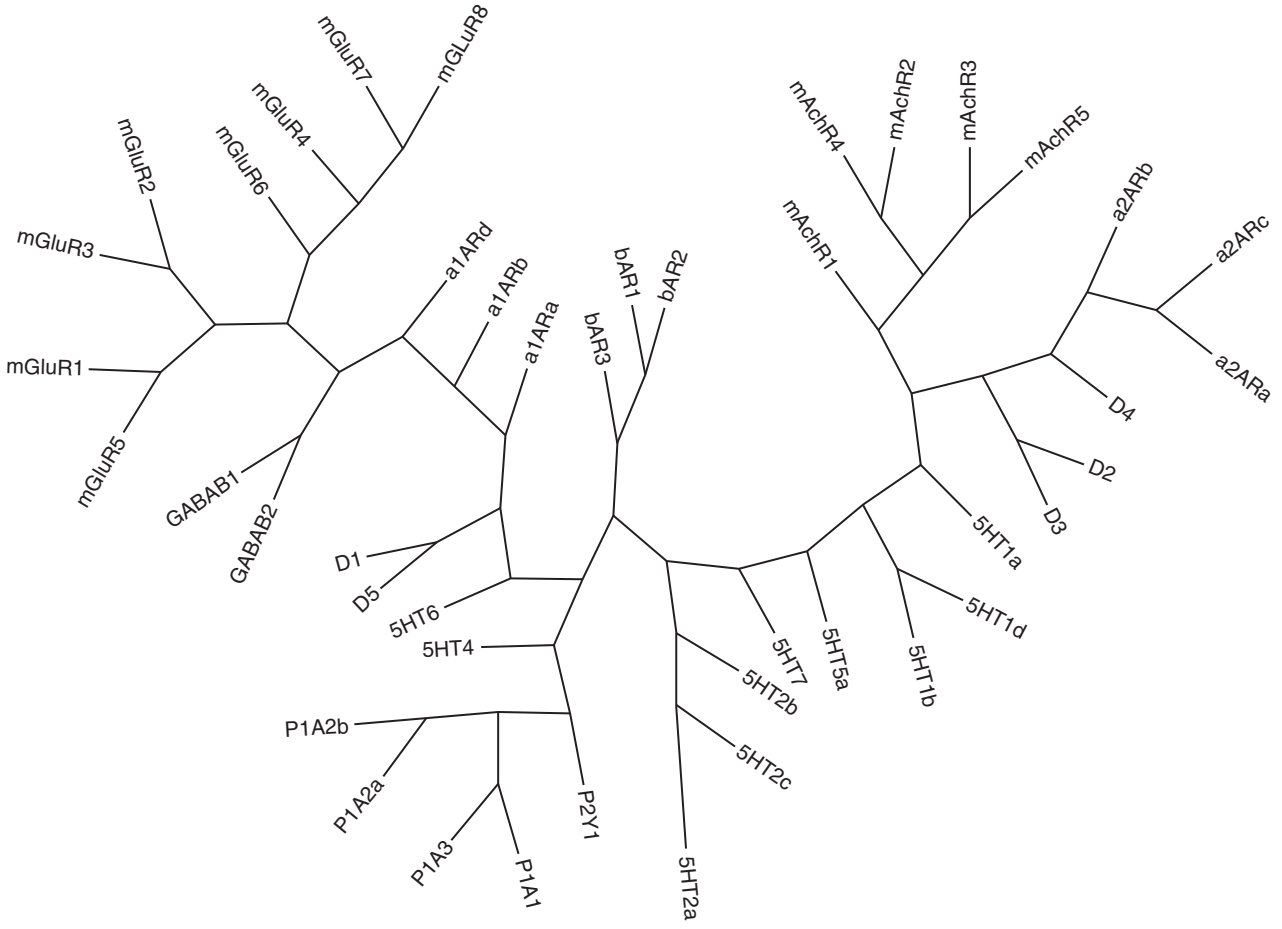


FIGURE 8.11 Evolutionary relationship of the GPCR family. The tree was constructed by aligning the protein sequences from each receptor family with the ClustalW program. Based on the alignment, the phylogenetic relationship was inferred with the maximum parsimony method and the tree was constructed using the Phylogeny Inference Package v 3.6 (distributed by J. Felsenstein, Department of Genome Sciences, University of Washington, Seattle.) Dr. Yin Liu (Department of Neurobiology and Anatomy, University of Texas Health Science Center-Houston, Houston, TX) kindly provided the phylogenetic tree and figure.

The family of mAChRs now includes five members (m1–m5), ranging from 55 to 70 kDa, and each of the five subtypes exhibits the typical architecture of seven TM domains. Much of the diversity in this family of receptors resides in the third intracellular loop (i3) responsible for the specificity of coupling to G-proteins. The m1, m3, and m5 mAChRs couple predominantly to G-proteins that activate the enzyme phospholipase C. m2 and m4 receptors (which are more highly related to each other than other family members; Fig. 8.11) couple to G-proteins that inhibit adenylyl cyclase, as well as to G-proteins that regulate K^+ and Ca^{2+} channels directly. As is the case for other GPCRs, the domain near the N terminus of i3 is important for the specificity of G-protein coupling. This domain is conserved in m1, m3, and m5 AChRs, but is unique in m2 and m4. Several other important residues also have been identified for G-protein coupling. A particular Asp residue near the N-terminus of the second intracellular loop (i2) is important for G-protein coupling, as are residues residing in the C-terminal region of the i3 loop.

Major mAChRs found in the brain are m1, m3, and m4, and each is distributed diffusely. The m2 subtype is the heart isoform and is not highly expressed in other organs. Genes for m4 and m5 lack introns, whereas those encoding m1, m2, and m3 contain introns, although little is known concerning alternatively spliced products of these receptors. Atropine is the most widely utilized antagonist for mAChR and binds to most subtypes, as does *N*-methylscopolamine. The antagonist pirenzipine appears to be relatively specific for the m1 mAChR, and other antagonists, such as AF-DX116 and hexahydrosiladifenidol, appear to be more selective for m2 and m3 subtypes.

Adrenergic Receptors

The catecholamines epinephrine (adrenaline) and norepinephrine (noradrenaline) produce their effects by

binding to and activating adrenergic receptors. Interestingly, epinephrine and norepinephrine can both bind to the same adrenergic receptor. Adrenergic receptors are currently separated into three families: α1, α2, and β (Fig. 8.11). Each of the α1 and α2 families is further subdivided into three subclasses. Similarly, the β family also contains three subclasses (β1, β2, and β3; Fig. 8.11). The main adrenergic receptors in the brain are the α1 and β1 subtypes. α2ARs have diverse roles, but the function that is best characterized (in both central and peripheral nervous tissue) is their role as autoreceptors. Different AR subtypes bind to G-proteins that can alter the activity of phospholipase C, Ca^{2+} channels, and, probably the best studied, adenylyl cyclase. For example, activation of α2ARs produces inhibition of adenylyl cyclase, whereas all βARs activate the cyclase.

Only a few agonists or antagonists cleanly distinguish the AR subtypes. One of them, isoproterenol, is an agonist that appears to be highly specific for βARs. Propranolol is the best-known antagonist for βARs, and phentolamine is a good antagonist for α receptors but binds weakly at βARs. The genomic organization of the different AR subtypes is unusual. Like many G-protein-coupled GPCRs, β1 and β2ARs are encoded by genes lacking introns. β3ARs, which have a role in lipolysis, are encoded by an intron-containing gene, as are αARs, providing an opportunity for alternative splicing as a means of introducing functional heterogeneity into the receptor.

Dopamine Receptors

Some 80% of the DA in the brain is localized to the corpus striatum, which receives major input from the substantia nigra and takes part in coordinating motor movements and other important information processing functions. DA is also found diffusely throughout the cortex, where its specific functions remain less well understood, but certainly play roles in higher cognitive functions. Many of the prescribed neuroleptic drugs exert their effects by blocking DA binding, and imbalances in the DA system have long been associated with neuropsychiatric disorders.

DA receptors are found both pre- and postsynaptically, and their structure is homologous to that of the receptors for other catecholamines (Civelli, Bunzow, & Grandy, 1993). Five subtypes of DA receptors can be grouped into two main classes: D1-like and D2-like receptors that are evolutionarily quite distinct (Fig. 8.11). D1-like receptors include D1 and D5, whereas D2-like receptors include D2, D3, and D4. The main distinction between these two classes is that D1-like receptors activate adenylyl cyclase through interactions with G_s, whereas D2-like receptors inhibit adenylyl cyclase and other effector molecules by interacting with G_i/G_o. D1-like receptors are also slightly larger in molecular mass than D2-like receptors. An additional point of interest is that D5 receptors selectively associate with $GABA_A$ receptors, impacting their function and vice versa (Liu et al., 2000). The deduced amino acid sequence for the entire family ranges from 387 amino acids (D4) to 477 amino acids (D5). Main structural differences between D1-like and D2-like receptors are that the intracellular loop between the sixth and the seventh transmembrane segments is larger in D2-like receptors, and D2-like receptors have smaller C-terminal intracellular segments.

D1-like receptors, like βARs, are transcribed from intronless genes. Conversely, all D2-like receptors contain introns, thus providing for possibilities of alternatively spliced products. Post-translational modifications include glycosylation at one or more sites, disulfide bonding of the two Cys residues in e2 and e3, and acylation of the Cys residue in the C-terminal tail (analogous to the β2AR). The DA-binding site includes two Ser residues in TM5 and an Asp residue in TM3, analogous to the βAR. The recent crystallization of the D3 receptor (Chien et al., 2010) supports the importance of these amino acid interactions with ligand, confirms the similarity to the β2AR, and provides the potential for future subtype specific drug design.

Because of the presumed role of DA in neuropsychiatric disorders, enormous effort has been put into developing pharmacological tools for manipulating this system. DA receptors bind bromocriptine, lisuride, clozapine, melperone, fluperlapine, and haloperidol. Because these drugs do not show great specificity for receptor subtypes, their usefulness for dissecting effects specifically related to binding to one or another DA receptor subtype is limited. However, their role in the treatment of human neuropsychiatric disorders is enormous (see Chapter 41).

Purinergic Receptors

Purinergic receptors bind to ATP or other nucleotide analogs and to its breakdown product adenosine. Although ATP is a common constituent found within synaptic vesicles, adenosine is not and is therefore not considered a "classic" neurotransmitter. However, the multitude of receptors that are activated by adenosine indicates that this molecule has important modulatory effects on the nervous system. Situations of high metabolic activity that consume ATP and situations of insufficient ATP-regenerating capacity can lead to the accumulation of adenosine. Because adenosine is permeable to membranes and can diffuse into and out of cells, a feedback loop is established in which adenosine can serve as a local diffusible signal that communicates the metabolic status of the neuron to surrounding cells and vice versa (Junger, 2011; Linden, 1994). In addition, all mammalian cells express extracellular ectonucleotidases that rapidly metabolize nucleotides into adenosine

providing an additional extracellular source of transmitter for the activation of adenosine receptors.

The original nomenclature describing purinergic receptors defined adenosine as binding to P1 receptors and ATP as binding to P2 receptors. Families of both P1 and P2 receptors have since been described, and adenosine receptors are now identified as A-type purinergic receptors, consisting of P1A1, P1A2a, P1A2b, and P1A3. ATP receptors are designated as P type and consist of P2X and P2Y subtypes (Junger, 2011). Recall that P2X subtypes are ionotropic receptors (see earlier discussion).

A-type receptors (P1A) exhibit the classic arrangement of seven transmembrane-spanning segments but are typically shorter than most GPCRs, ranging in size between 35 and 46 kDa. The ligand-binding site of A-type receptors is unique in that the ligand, adenosine, has no inherent charged moieties at physiological pH. A-type receptors appear to utilize His residues as their points of contact with adenosine, and, in particular, a His residue in TM7 is essential because its mutation eliminates agonist binding. Other His residues in TM6 and TM7 are conserved in all A-type receptors and may serve as other points of contact with agonists. A1 receptors are highly expressed in the brain, and their activation down regulates adenylyl cyclase and increases phospholipase C activity. The A2a and A2b receptors are not as highly expressed in nervous tissue and are associated with the stimulation of adenylyl cyclase and phospholipase C, respectively. The A3 subtype exhibits a unique pharmacological profile in that binding of xanthine derivatives (caffeine falls in this class of molecules), which blocks the action of adenosine competitively, is much weaker than for A1 or A2 receptors. Very low levels of the A3 receptor are found in brain and peripheral nervous tissue. The A3 receptor appears to be coupled to the activation of phospholipase C.

The P-type receptors, P2Y1–P2Y7, are also typical G-protein-linked GPCRs, mostly localized to the periphery. However, direct effects of ATP have been detected in neurons, and often the response is biphasic, an early excitatory effect followed, with its breakdown to adenosine, by a secondary inhibitory effect. Interestingly, P-type receptors exhibit a higher degree of homology to peptide-binding receptors than they do to A-type purinergic receptors. As in A-type receptors, P-type receptors have a His residue in the third transmembrane domain and presumably bind to nucleotides using a similar constellation of amino acids as other members of the class A GPCRs.

Serotonin Receptors

Cell bodies containing serotonin (5-HT) are found in the raphe nucleus in the brain stem and in nerve endings distributed diffusely throughout the brain. 5-HT has been implicated in sleep, modulation of circadian rhythms, eating, and arousal. 5-HT also has hormone-like effects when released in the bloodstream, regulating smooth muscle contraction and affecting platelet-aggregating and immune systems.

5-HT receptors are classified into seven subtypes, 5-HT1 to 5-HT7 (Fig. 8.11), with a further subdivision within subtypes. Recall that the 5-HT3 receptor is ionotropic (see earlier discussion). The other 5-HT receptors exhibit the typical seven transmembrane-spanning segments and all couple to G-proteins to exert their effects. For example, 5-HT1a, 1b, 1d, and 4, 5, 6, and 7 either activate or inhibit adenylyl cyclase. 5-HT1c and 5-HT2 receptors preferentially stimulate activation of phospholipase C to produce increased intracellular levels of diacylglycerol and inositol 1,4,5-trisphosphate.

5-HT receptors can also be grossly distributed into two groups on the basis of their gene structures. Both 5-HT1c and 5-HT2 are derived from genes that contain multiple introns. In contrast, similar to the βAR family, 5-HT1 is coded by a gene lacking introns. Interestingly, 5-HT1a is more closely related ancestrally to the D2 receptor family than it is to other membranes of the 5-HT receptor family that are separated quite far apart on the evolutionary tree (Fig. 8.11).

Glutamate GPCRs

GPCRs that bind glutamate (metabotropic glutamate receptors (mGluRs)) are similar in general structure in having seven transmembrane-spanning segments to other GPCRs; however, they are divergent enough to be placed in a separate evolutionary-derived receptor family (Fig. 8.11; Hollmann & Heinemann, 1994; Nakanishi, 1994; Niswender & Conn, 2010). In fact, sequence homology between the mGluR family and other GPCRs is minimal except for the $GABA_B$ receptor (Fig. 8.11) and these two receptor families along with Ca^{2+}-sensing and taste receptors are now grouped together as class C GPCRs. The mGluR family is heterogeneous in size, ranging from 854 to 1179 amino acids. Both the N-terminal and the C-terminal domains are unusually large for G-protein-coupled receptors. The binding site for glutamate resides in the large N-terminal extracellular domain and several crystal structures of this domain in different states have been obtained. The binding site is called the Venus Fly trap Domain (VFD) to anecdotally capture the structural transition of the open to closed nature of the domain when bound to glutamate. Recall that in Class A GPCRs, the ligand-binding pocket sits within the TM domains of the receptor. Additionally, mGluRs exist as functional dimers in the membrane in contrast to the single subunit forms of most GPCRs (Kunishima et al., 2000; Niswender & Conn, 2010).

Because of this dimer structure, the mGluRs can exist in states with either one or two molecules of glutamate bound. Maximal activity requires that two glutamate molecules are bound, but there remains some controversy about whether a single molecule of glutamate can activate the receptor. The second intracellular loop (i2), as opposed to the third intracellular loop in class A GPCRs, along with the large C-terminal domain are thought to be the major determinants responsible for G-protein coupling. The C-terminal domain also serves as the site for interactions with other intracellular scaffolding and regulatory molecules. These significant structural distinctions support the idea that mGluRs evolved separately from other GPCRs.

Currently, eight different mGluRs can be subdivided into three groups on the basis of sequence homologies (Fig. 8.11) and their capacity to couple to specific enzyme systems. Group I includes mGluR1 and mGluR5, and these receptors are found predominantly postsynaptic and activate G-proteins that stimulate phospholipase C or adenylyl cyclase. Group II includes mGluR2 and mGluR3 that are found pre- and postsynaptic, inhibit adenylyl cyclase, inhibit Ca^{2+}-channels, and activate K^+-channels. Group III includes mGluR4–mGluR8 and produce functional effects similar to Group II mGluRs. mGluR6 is unique and is found only in the retina, where it stimulates the enzyme cGMP phosphodiesterase.

Group I mGluRs are selectively activated by the agonist (S)-3,5-dihydroxyphenylglycine [(S)-3,5-DHPG)]. Group II mGluRs are selectively activated by (2S,2′R,3′R)-2-(2′,3′)-Dicarboxycyclopropyl)glycine DCG-IV and (2R,4R)-4-Aminopyrrolidine-2,4-dicarboxylate (2R,4R-APDC) and group II mGluRs are selectively activated by L-(+)-2-Amino-4-phosphonobutyric acid (L-AP4). Numerous other pharmacological compounds have also been developed that can selectively activate members within some of these mGluR classes.

mGluRs are widespread in the nervous system and are found both pre- and postsynaptically on neurons and are also abundant on glia. Generalizations about the specific functions of each class or subtype are difficult. Presynaptically, they serve as autoreceptors and appear to participate in the inhibition of neurotransmitter release. Their postsynaptic roles appear to be quite varied and depend on the specific G-protein to which they are coupled. mGluR1 activation has been implicated in long-term synaptic plasticity at many sites in the brain, including long-term potentiation in the hippocampus and long-term depression in the cerebellum (see Chapter 47).

$GABA_B$ Receptor

$GABA_B$ receptors are found throughout the nervous system, where they are sometimes colocalized with ionotropic $GABA_A$ receptors. $GABA_B$ receptors are present both pre- and postsynaptically. Presynaptically, they appear to mediate inhibition of neurotransmitter release through an autoreceptor-like mechanism by activating K^+ conductances and diminishing Ca^{2+} conductances. In addition, $GABA_B$ receptors may affect K^+ channels through a direct physical coupling to the K^+ channel, not mediated through a G-protein intermediate. Postsynaptically, $GABA_B$ receptor activation produces a characteristic slow hyperpolarization (termed the slow inhibitory postsynaptic potential) through the activation of a K^+ conductance. This effect appears to be through a pertussis toxin-sensitive G-protein that inhibits adenylyl cyclase. The initial cloning of the $GABA_B$ receptor was made possible by the development of the high-affinity, high-specificity antagonist CGP64213. This antagonist is several orders of magnitude more potent at inhibiting $GABA_B$ receptor function than the more widely known antagonist saclofen. Baclofen, an analog of saclofen, remains a common agonist for activating $GABA_B$ receptors.

Cloning of the $GABA_B$ receptor ($GABA_BR1$) revealed that it has high sequence homology to the family of glutamate GPCRs and is also a member of the class C GPCRs that shows little similarity to other GPCRs (Fig. 8.11). The large N-terminal extracellular domain of the $GABA_B$ receptor is the presumed site of GABA binding. With the exception of this large extracellular domain, the $GABA_B$ receptor structure is typical of the GPCR family, exhibiting seven TM domains. Functional $GABA_B$ receptors exist as obligate dimers in the membrane. Expression of the $GABA_B1$ subunit does not produce functional receptors, but when coexpressed with the $GABA_B2$ subunit, receptors that are indistinguishable functionally and pharmacologically from those in brain were produced. GABA binds to the $GABA_B1$ subunit to activate the receptor and the $GABA_B2$ subunit is responsible for coupling to G-proteins (Kniazeff, Prezeau, Rondard, Pin, & Goudet, 2011). As noted earlier, GPCRs can interact with themselves and other receptors. It is well to keep in mind that these types of direct receptor interactions may be more widespread than currently appreciated.

Peptide Receptors

Neuropeptide receptors form an immense family and make up the majority of Class B GPCRs. Because of their diversity, they cannot be covered in detail in this chapter. Despite this diversity, however, none of the receptors that bind peptides appears to be ionotropic (coupled directly to the opening of ion channels). Neuropeptide receptors exert their effects either through the typical pathway of activation of G-proteins or through the activation of an associated tyrosine kinase activity.

Summary

GPCRs are typically single polypeptides composed of seven TM-spanning segments. The binding site for neurotransmitter in Class A GPCRs is located within the core of the circular structure formed by these TM segments. Transmitter binding produces conformational changes in the receptor that expose parts of the i3 region, among others, for binding to G-proteins that activate additional downstream enzymes or other effectors. G-protein binding increases the affinity of the receptor for transmitter. Desensitization is common among GPCRs and leads to a decreased response of the receptor to neurotransmitter by several distinct mechanisms. Class C GPCRs are structurally distinct from the other classes. They have large N-terminal extracellular domains that form the binding site for neurotransmitter. Additionally, functional class C GPCRs appear to function as homo- or heterodimers in the membrane, but otherwise couple to G-proteins in a similar manner to the rest of the GPCR family.

References

Boulter, J., Hollmann, M., O'Shea-Greenfield, A., Hartley, M., Deneris, E., Maron, C., & Heinemann, S. (1990). Molecular cloning and functional expression of glutamate receptor subunit genes. *Science, 249,* 1033–1037.

Burnashev, N., Schoepfer, R., Monyer, H., Ruppersberg, J. P., Gunther, W., Seeburg, P. H., & Sakmann, B. (1992). Control by asparagine residues of calcium permeability and magnesium blockade in the NMDA receptor. *Science, 257*, 1415-1419.

Chien, E. Y., Liu, W., Zhao, Q., Katritch, V., Han, G. W., Hanson, M. A., Shi, L., Newman, A. H., Javitch, J. A., Cherezov, V., & Stevens, R. C. (2010). Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science, 330,* 1091–1095.

Civelli, O., Bunzow, J. R., & Grandy, D. K. (1993). Molecular diversity of the dopamine receptors. *Annual Review of Pharmacology and Toxicology, 33,* 281–307.

Ferguson, S. S. G., Downey, W. E., Colapietro, A.-M., Barak, L. S., Menard, L., & Caron, M. G. (1996). Role of arrestin in mediating agonist-promoted G-protein coupled receptor internalization. *Science, 271,* 363–366.

Hollmann, M., & Heinemann, S. (1994). Cloned glutamate receptors. *Annual Review of Neuroscience, 17,* 31–108.

Hollmann, M., O'Shea-Greenfield, A., Rogers, S. W., & Heinemann, S. (1989). Cloning by functional expression of a member of the glutamate receptor family. *Nature (London), 342,* 643–648.

Huang, C. -C., & Tesmer, J. J. G. (2011). Recognition in the face of diversity: interactions of heteromeric G protein and G protein-coupled receptor (GPCR) kinases with activated GPCRs. *Journal of Biological Chemistry, 268,* 7715–7721.

Junger, W. G. (2011). Immune cell regulation by autocrine purinergic signalling. *National Review Immunology, 11,* 201–212.

Keinanen, K., Wisden, W., Sommer, B., Werner, P., Herb, A., Verdoorn, T. A., Sakmann, B., & Seeburg, P. H. (1990). A family of AMPA-selective glutamate receptors. *Science, 249,* 556–560.

Kniazeff, J., Prezeau, L., Rondard, P., Pin, J. -P., & Goudet, C. (2011). Dimers and beyond: the functional puzzles of class C GPCRs. *Pharmacology and Therapeutics, 130,* 9–25.

Kobilka, B. (1992). Adrenergic receptors as models for G-protein-coupled receptors. *Annual Review of Neuroscience, 15,* 87–114.

Kunishima, N., Shimada, Y., Tsuji, Y., Toshihiro Sato, Masaki Yamamoto, Takashi Kumasaka, Shigetada Nakanishi, Hisato Jingami, & Kosuke Morikawa (2000). Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor. *Nature, 407,* 971–977.

Linden, J. (1994). In Siegel, G. J., Agranoff, B. W., Albers, R. W. & Molinoff, P. B. (Eds.), *Basic neurochemistry* (pp. 401–416). New York: Raven Press.

Liu, F., Wan, Q., Pristupa, Z. B., Yu, X. -M., Want, Y. T., & Niznik, H. B. (2000). Direct protein-protein coupling enables cross-talk between dopamine D5 and g-aminobutyric acid A receptors. *Nature, 403,* 274–278.

Nakagawa, N., Cheng, Y., Ramm, E., Sheng, M., & Walz, T. (2005). Structure and different conformational states of native AMPA receptor complexes. *Nature, 433,* 545–549.

Nakanishi, S. (1994). Metabotropic glutamate receptors: Synaptic transmission, modulation, and plasticity. *Neuron, 13,* 1031–1037.

Nakanishi, N., Shneider, N. A., & Axel, R. (1990). A family of glutamate receptor genes: Evidence for the formation of heteromultimeric receptors with distinct channel properties. *Neuron, 5,* 569–581.

Niswender, C. M., & Conn, P. J. (2010). Metabotropic glutamate receptors: physiology, pharmacology and disease. *Annual Review of Pharmacology and Toxicology, 50,* 295–322.

Premont, R. T., Inglese, J., & Lefkowitz, R. J. (1995). Protein kinases that phosphorylate activated G-protein-coupled receptors. *FASEB Journal, 9,* 175–182.

Rosenbaum, D. M., Rasmussen, S. G. F., & Kobilka, B. K. (2009). The structure and function of G-protein coupled receptors. *Nature (London), 459,* 356–363.

Sheng, M., & Hoogenraad, C. C. (2007). The postsynaptic architecture of excitatory synapses: A more quantitative view. *Annual Review of Biochemistry, 76,* 823–847.

Sobolevsky, A. I., Rosconi, M. P., & Gouaux, E. (2009). X-ray structure, symmetry and mechanism of an AMPA-subtype glutamate receptor. *Nature (London), 446,* 745–758.

Sommer, B., Kohler, M., Sprengel, R., & Seeburg, P. H. (1991). RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell (Cambridge, Mass), 67,* 11–19.

Strader, C. D., Fong, T. M., Tota, M. R., Underwood, D., & Dixon, R. A. (1994). Structure and function of G-protein-coupled receptors. *Annual Review of Biochemistry, 63,* 101–132.

Strosberg, A. D. (1990). Biotechnology of β-adrenergic receptors. *Molecular Neurobiology, 4,* 211–250.

Traynelis, S. F., Wollmuth, L. P., McBain, C. J., Menniti, F. S., Vance, K. M., Ogden, K. K., Hansen, K. B., Yuan, H., Myers, S. J., & Dingledine, R. (2010). Glutamate receptor ion channels: structure, regulation and function. *Pharmacological Reviewa, 62,* 405–496.

Unwin, N. (1995). Acetylcholine receptor channel imaged in the open state. *Nature (London), 373,* 37–43.

Unwin, N. (2005). Refined structure of the nicotinic acetylcholine receptor at 4 A resolution. *Journal of Molecular Biology, 346,* 967–989.

Verdoorn, T. A., Burnashev, N., Monyer, H., Seeburg, P. H., & Sakmann, B. (1991). Structural determinants of ion flow through recombinant glutamate receptor channels. *Science, 252,* 1715–1718.

Watkins, J. C., Krogsgaard-Larsen, P., & Honore, T. (1990). Structure activity relationships in the development of excitatory amino acid receptor agonists and competitive antagonists. *Trends in Pharmacological Sciences, 11,* 25–33.

Willmann, R., & Fuhrer, C. (2002). Neuromuscular synaptogenesis: clustering of acetylcholine receptors revisited. *Cellular and Molecular Life Sciences, 59,* 1296–1316.

Wo, Z. G., & Oswald, R. E. (1995). Unraveling the modular design of glutamate-gated ion channels. *Trends in Neurosciences, 18,* 161–168.

M. Neal Waxham

This page intentionally left blank

CHAPTER

9

Intracellular Signaling

SIGNALING THROUGH G-PROTEIN-LINKED RECEPTORS

Signal transduction through G-protein-linked receptors requires three membrane-bound components: (1) a cell surface receptor that determines to which signal the cell can respond; (2) a G protein on the intracellular side of the membrane that is stimulated by the activated receptor; and (3) either an effector enzyme that changes the level of a second messenger or an effector channel that changes ionic fluxes in the cell in response to the activated G protein. The human genome encodes for more than 800 receptors for catecholamines, odorants, neuropeptides, and light that couple to one or more of the 16 identified G proteins. These, in turn, regulate one or more of more than two dozen different effector channels and enzymes. The key feature of this information flow is the ability of G proteins to detect the presence of activated receptors and to amplify the signal by altering the activity of appropriate effector enzymes and channels. A nervous system with information flow by fast transmission alone would be capable of stereotyped or reflex responses. Modulation of this transmission and changes in other cellular functions by G-protein-linked systems and by receptor-tyrosine kinase-linked systems enables an orchestrated response. The large diversity of signaling molecules and their intracellular targets offer nearly unlimited flexibility of response over a broad time scale and with high amplification.

G proteins are GTP-binding proteins that couple the activation of seven-helix receptors by neurotransmitters at the cell surface to changes in the activity of effector enzymes and effector channels. A common effector enzyme is adenylate cyclase, which synthesizes cyclic AMP (cAMP)—an intracellular surrogate for the neurotransmitter, the first messenger. Phospholipase C (PLC), another effector enzyme, generates diacylglycerol (DAG) and inositol 1, 4, 5-trisphosphate (IP_3), the latter of which releases intracellular stores of Ca^{2+}. Information from an activated receptor flows to the second messengers that typically activate protein kinases, which modify a host of cellular functions. Ca^{2+}, cAMP, and DAG have in common the ability to activate protein kinases with broad substrate specificities. They phosphorylate key intracellular proteins, ion channels, enzymes, and transcription factors taking part in diverse cellular biological processes. The activities of protein kinases and phosphatases are in balance, constituting a highly regulated process, as revealed by the phosphorylation state of these targets of the signal transduction process. In addition to regulating protein kinases, second messengers such as cAMP, cyclic GMP (cGMP), Ca^{2+}, and arachidonic acid can directly gate, or modulate, ion channels. G proteins can also couple directly to ion channels without the interception of second messengers or protein kinases. In these diverse ways, a neurotransmitter outside the cell can modulate essentially every aspect of cell physiology and encode the history of cell stimuli in the form of altered activity and expression of its cellular constituents. An overview of G-protein signaling to protein kinases is presented in Fig. 9.1.

Receptors Catalyze the Conversion of G Proteins into the Active GTP-Bound State

G proteins undergo a molecular switch between two interconvertible states that are used to "turn on" or "turn off" downstream signaling. G proteins taking part in signal transduction utilize a regulatory motif that is seen in other GTPases engaged in protein synthesis and in intracellular vesicular traffic. G proteins are switched on by stimulated receptors, and they switch themselves off after a time delay. G proteins are inactive when GDP is bound and are active when GTP is bound. The sole function of seven-helix receptors in activating

© 2013 Elsevier Inc. All rights reserved.

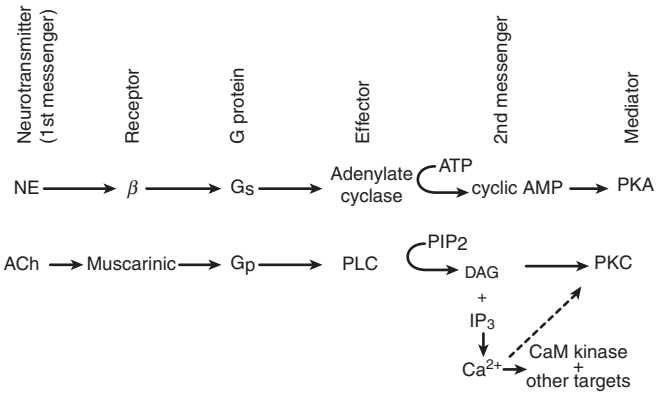


FIGURE 9.1 Overview of G-protein signaling to protein kinases. Norepinephrine (NE) and acetylcholine (ACh) can stimulate certain receptors that couple through distinct G proteins to different effectors, which results in increased synthesis of second messengers and activation of protein kinases (PKA and PKC). PLC, phospholipase C; PIP_2, phosphatidylinositol bisphosphate; DAG, diacylglycerol; CaM, Ca^{2+}-calmodulin dependent; IP_3, inositol 1, 4, 5-triphosphate.

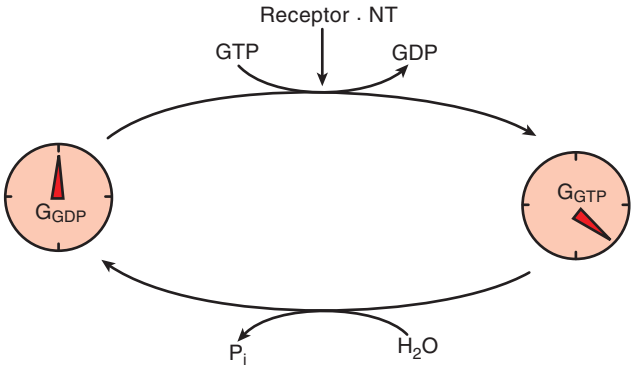


FIGURE 9.2 GTPase activity of G proteins serves as a timer and amplifier. Receptors activated by neurotransmitters (NT) initiate the GTPase timing mechanism of G proteins by displacement of GDP by GTP. Neurotransmitters thus convert G-GDP ("turned-off state") to G-GTP (time-limited "turned-on" state).

G proteins is to catalyze an exchange of GTP for GDP. This is a temporary switch because G proteins are designed with a GTPase activity that hydrolyzes the bound GTP and converts the G protein back into the GDP-bound, or inactive, state. Thus, a G protein must continuously sample the state of activation of the receptor, and it transmits downstream information only while the neuron or glial cell is exposed to neurotransmitter. The GTPase activity of G proteins thus serves both as a regulatable timer and as an amplifier (Fig. 9.2).

The G-Protein Cycle

G proteins are trimeric structures composed of two functional units: (1) an α subunit (39–52 kDa) that

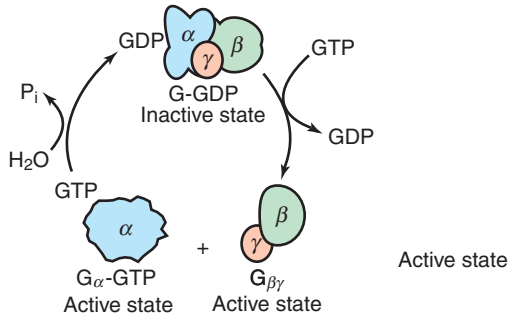


FIGURE 9.3 Interconversion, catalyzed by excited receptors, of G-protein subunits between inactive and active states. Displacement of GDP with GTP dissociates the inactive heterotrimeric G protein, generating α-GTP and $\beta\gamma$, both of which can interact with their respective effectors and activate them. The system converts into the inactive state after GTP has been hydrolyzed and the subunits have reassociated. *From Stryer (1995). Used with permission of W. H. Freeman and Company.*

catalyzes GTPase activity and (2) a $\beta\gamma$ dimer (35 and 8 kDa, respectively) that interacts tightly with the α subunit when bound to GDP (Stryer & Bourne, 1986; Birnbaumer, 2007). The role of the three subunits in the G-protein cycle is depicted in Figures 9.3 and 9.4. In the basal state, GDP is bound tightly to the α subunit, which is associated with the $\beta\gamma$ pair to form an inactive G protein. In addition to blocking interaction of the α subunit with its effector, the $\beta\gamma$ pair increases the affinity of the α subunit for activated receptors. Binding of the neurotransmitter to the receptor produces a conformational change that positions previously buried residues that promote increased affinity of the receptors for the inactive G protein. A given receptor can interact with only one or a limited number of G proteins, and the α subunit generates most of this specificity. Coupling with the activated receptor reduces the affinity of the α subunit for GDP, facilitating its dissociation and replacement with GTP. Thus, the receptor effectively catalyzes an exchange of GTP for GDP. GTP-GDP exchange is inherently very slow and ensures that very little of the G protein is in the on state under basal conditions. The level of G protein in the "on" state can increase from being 1% to being more than 50% of all G protein (Stryer & Bourne, 1986).

Information Flow through G-Protein Subunits

One of the more tense and public debates in signal transduction has been the question of whether the α subunit alone conveys information that specifies which effector is activated or whether the $\beta\gamma$ pair can also interact with effectors. One of the contestants even paid for a vanity license plate proclaiming "α not $\beta\gamma$." This notion was eventually changed because of the finding that $\beta\gamma$ can directly activate certain K^+ channels. It is now apparent that α and $\beta\gamma$ subunits can both modify effector

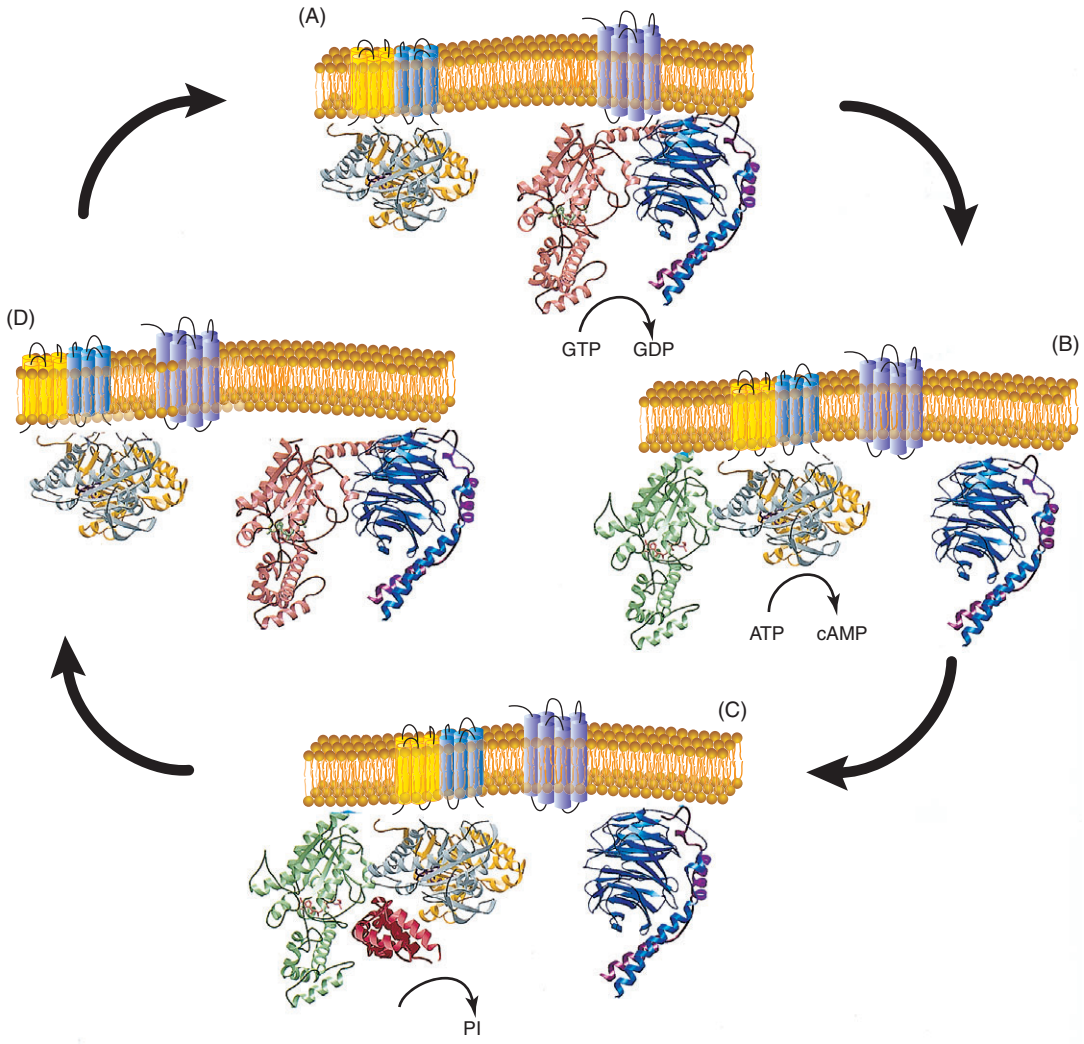


FIGURE 9.4 (A) G proteins are held in an inactive state because of very high affinity binding of GDP to their α subunits. When activated by agonist, membrane-bound seven helical receptors (Fig. 9.4 right, glowing magenta) interact with heterotrimeric G proteins (α, amber; β, teal; γ, burgundy) and stimulate dissociation of GDP. This permits GTP to bind to and activate α, which then dissociates from the high-affinity dimer of β and γ subunits. (B) Both activated (GTP-bound) α (lime) and $\beta\gamma$ are capable of interacting with downstream effectors. Figure 9.4 shows the interaction of GTP-α_s with adenylate cyclase (catalytic domains are mustard and ash). Adenylate cyclase then catalyzes the synthesis of the second messenger cyclic AMP (cAMP) from ATP. (C) Signaling is terminated when α hydrolyzes its bound GTP to GDP. In some signaling systems, GTP hydrolysis is stimulated by GTPase-activating proteins or GAPs (cranberry) that bind to α and stablize the transition state for GTP hydrolysis. (D) Hydrolysis of GTP permits GDP-α to dissociate from its effector and associate again with $\beta\gamma$. The heterotrimeric G protein is then ready for another signaling cycle if an activated receptor is present. This figure is based on the original work of Mark Wall and John Tesmer.

enzymes, but the historic association of G-protein function with α has persisted for the purpose of nomenclature, with G_s and α_s referring to the G protein and its corresponding α subunit, which stimulates adenylate cyclase. The α subunits may act either independently or in concert with $\beta\gamma$ (Clapham & Neer, 1993). Furthermore, β and γ subunits in a $\beta\gamma$ pair can combine in many different ways. Other legacy terms include G_i, G_p, and G_o used for G-protein activities that inhibited adenylate cyclase, stimulated phospholipase, or were presumed to have other effects, respectively.

Effector Enzymes, Channels, and Transporters Decode Receptor-Mediated Cell Stimulation in the Cell Interior

The function of the trimeric G proteins is to decode information about the concentration of neurotransmitters bound to appropriate receptors on the cell surface and convert this information into a change in the activity of enzymes and channels that mediate the effects of the neurotransmitter. The known effector functions of α include both stimulation and inhibition of adenylate

cyclases that is sensitive to cholera toxin and pertussis toxin, respectively. In addition, it modulates activation of cGMP phosphodiesterase, PLC, and regulation of Na^+/K^+ exchange, PI3K, RhoGEF and rasGAP. The effector functions of $\beta\gamma$ dimers include inhibition of many adenylate cyclases and co-stimulation of others. In addition, they regulate stimulation of phospholipase Cβ, K^+ and Ca^{2+} channels, phospholipase A_2, phosphatidylinositol-3-kinase, PKD and dynamin in vesicle budding.

Response Specificity in G-Protein Signaling

How can a neurotransmitter produce a specific response if G-protein coupling has the potential for such a diversity of effectors? A given neuron has only a subset of receptors, G proteins, and effectors, thereby limiting possible signaling pathways. Transducin, for example, is confined to the visual system, where the predominant effector is the cGMP phosphodiesterase and not adenylate cyclase. Signal specificity is further refined by selective affinities between cognate sets of receptors, G proteins, and effector(s), and by spatial compartmentalization, such as at nerve terminals, where co-localization effects privileged signaling from a given receptor to a limited set of effectors. Furthermore, the intrinsic GTPase activity of G proteins can be modulated by GTPase-activating proteins (GAPs) that terminate its active state more quickly and selectively affect signal output.

Fine-Tuning of cAMP by Adenylate Cyclases

The level of cAMP is highly regulated due to a balance between synthesis by adenylate cyclases and degradation by cAMP phosphodiesterases (PDEs). Each of these enzymes can be regulated and manipulated independently. Adenylate cyclase was the first G-protein effector to be identified, and now a group of related adenylate cyclases are known to be regulated differentially by both α and $\beta\gamma$ subunits (Taussig & Gilman, 1995). G proteins can both activate and inhibit adenylate cyclases either synergistically or antagonistically.

Adenylate cyclases are large proteins of approximately 120 kDa. They consist of a tandem repeat of the same structural motif—a short cytoplasmic region followed by six putative transmembrane segments and then a highly conserved catalytic domain of approximately 35 kDa on the cytoplasmic side that bind ATP and catalyze its conversion into cAMP. Some isoforms are activated by calmodulin, a Ca^{2+}-binding protein that activates many enzymes after it binds Ca^{2+}.

Differential Regulation of Adenylate Cyclase Isoforms

All adenylate cyclase isoforms are stimulated by G_s through its α_s subunit. Known isoforms can be divided into four groups on the basis of additional regulatory properties (Fig. 9.5). Group I (AC1, 3, 8) possesses a calmodulin-binding domain and is activated by Ca^{2+}-calmodulin. Group II (AC2, 4, 7) is weakly responsive to direct interaction with α_s or $\beta\gamma$ but is highly activated when both are present. As described later, this synergistic effect enables this cyclase to function as a coincidence detector. Group III cyclases (AC5, 6) are negatively regulated by inhibitory G proteins, whereas Group IV (AC9) has a forskolin-insensitive cyclase.

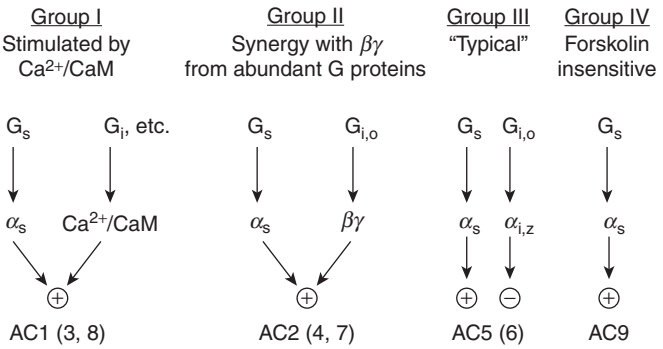


FIGURE 9.5 Isoforms of adenylate cyclase (AC). All isoforms are stimulated by α_s but differ in the degree of interaction with Ca^{2+}-calmodulin and with $\beta\gamma$ derived from inhibitory G proteins. Not shown is the ability of excess $\beta\gamma$ to complex with α_s and inhibit Group I and Group III adenylate cyclases. *Updated from Taussig and Gilman (1995).*

INHIBITION OF ADENYLATE CYCLASES

Adenylate cyclases are also subject to several forms of inhibitory control. First, activation of all adenylate cyclases can be antagonized by $\beta\gamma$ released from abundant G proteins, such as G_i, G_o, and G_z. Released $\beta\gamma$ complexes with α_s-GTP and shift the equilibrium toward an inactive trimer by mass action. Second, either α or $\beta\gamma$ subunits derived from G_i, G_o, or G_z can directly inhibit group I cyclases, and the α subunit from G_i or G_z can inhibit Group III cyclases. The level of G_s in particular is low; thus, a_s derived from G_s is sufficient to activate adenylate cyclases, but the $\beta\gamma$ derived from it is insufficient to directly inhibit or activate adenylate cyclases. This explains the apparent paradox that receptors that couple to G_s produce effects only through a_s, whereas receptors that couple to G_i produce effects through both α_i and $\beta\gamma$. Additional inhibitory regulation of some groups is also effected by CaM kinases and by calcium.

RECEPTORS COUPLING TO ADENYLATE CYCLASE

Dozens of neurotransmitters and neuropeptides work through cAMP as a second messenger and by G protein-linked activation or inhibition of adenylate cyclase. Among the neurotransmitters that increase cAMP are the amines norepinephrine, epinephrine, dopamine, serotonin, and histamine and the neuropeptides vasointestinal peptide (VIP) and somatostatin. In the olfactory system, a special form of G-protein α subunit, termed

α_{olf}, serves the same function as α_s and couples several hundred seven-helix receptors to AC3 adenylate cyclase in the neuroepithelium.

ADENYLATE CYCLASES AS COINCIDENCE DETECTORS

Adenylate cyclases can integrate concurrent stimulation of neurons by two or more neurotransmitters (Bourne & Nicoll, 1993). AC1 adenylate cyclase is stimulated by neurotransmitters that couple to G_s and by neurotransmitters that elevate intracellular Ca^{2+}. This adenylate cyclase can convert the depolarization of neurons into an increase in cAMP. Its role in associative forms of learning may be related to its ability to link cAMP-based and Ca^{2+}-based signals. Stimulation of AC2 adenylate cyclase by α_s is conditional on the presence of $\beta\gamma$ derived from an abundant G protein—that is, other than G_s, thus enabling the cyclase to serve as a coincidence detector. Thus, activation of a second receptor, presumably coupled to the abundant G_i and G_o, is needed to provide the $\beta\gamma$.

Sources of Second Messengers: Phospholipids

Two phospholipids—phosphatidylinositol 4,5-bisphosphate (PIP_2) and phosphatidylcholine (PC)—are primary precursors for a G-protein-based second-messenger system. Three second-messengers—diacylglycerol, arachidonic acid and its metabolites, and elevated Ca^{2+}—are ultimately produced. A single step converts inert phospholipid precursors into lipid messengers. DAG action is primarily mediated by protein kinase C (PKC) (Tanaka & Nishizuka, 1994). The elevation of Ca^{2+} levels is accomplished by the regulated entry of Ca^{2+} from a concentrated pool sequestered in the endoplasmic reticulum or from outside the cell. Ca^{2+} has many direct cellular targets but mediates most of its effects through Ca^{2+}/calmodulin. One class of calmodulin-dependent enzymes is a family of protein kinases that enable Ca^{2+} signals to regulate a large number of cellular processes by phosphorylation (Wayman et al., 2008).

GENERATION OF DAG AND IP_3 FROM G_q AND G_i COUPLED TO PLCβ

The phosphatidylinositide-signaling pathway is just as prominent in neuronal signaling as the cAMP pathway and is similar to it in overall design (Berridge, 2009). Stimulation of a large number of neurotransmitters and hormones [including acetylcholine (Ml, M3), serotonin ($5HT_2$, $5HT_{1C}$), norepinephrine (α_{1A}, α_{1B}), glutamate (metabotropic), neurotensin, neuropeptide Y, and substance P] is coupled to the activation of a phosphatidylinositide-specific PLC.

Phosphatidylinositol (PI) is composed of a diacylglycerol backbone with myoinositol attached to the *sn*-3 hydroxyl by a phosphodiester bond (Fig. 9.6). The *six* positions of the inositol are not equivalent: the 1 position is attached by a phosphate to the DAG moiety. PI is phosphorylated by PI kinases at the 4 position and then at the 5 position to form PIP_2. In response to the appropriate G-protein coupling, PLC hydrolyzes the bond between the *sn*-3 hydroxyl of the DAG backbone and the phosphoinositol to produce two second messengers: DAG, a hydrophobic molecule, and IP_3, which is water soluble (Fig. 9.7). There are six classes of PLCs that hydrolyze PIP_2 (PLCβ, γ, δ, ε, ζ, and η), with a common catalytic domain structure but different regulatory properties. G proteins couple to several variants of PLCβ. PLCγ is regulated by growth factor tyrosine kinases. Less is known about regulation of the others, but PKCε can be regulated by both G proteins and small GTPases.

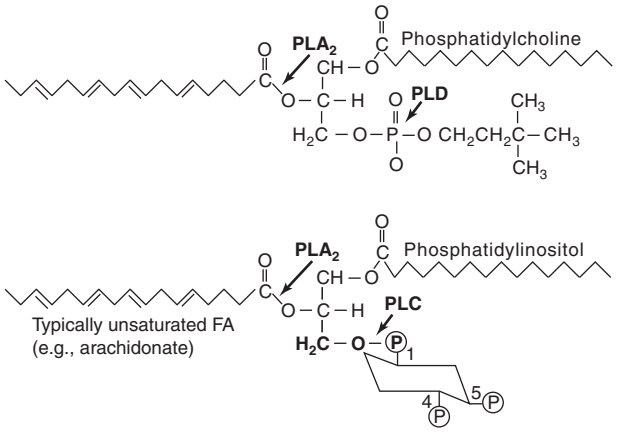


FIGURE 9.6 Structures of phosphatidylinositol and phosphatidylcholine. The sites of hydrolytic cleavage by PLC, PLD, and PLA_2 are indicated by arrows. FA, fatty acid.

PLCβ is coupled to neurotransmitters by G_i and G_q. A pertussis toxin-sensitive pathway is mediated by a number of isoforms referred to as G_q and mediated by their α_q. G_i is coupled to PLCβ via its $\beta\gamma$ rather than α subunit in a pathway that is insensitive to pertussis toxin. Receptor tyrosine kinases can regulate PLCγ by a G-protein-independent pathway involving their recruitment to the receptor and activation via phosphorylation.

DAG DERIVED FROM ACTIVATION OF PHOSPHOLIPASE D

A slower but larger increase in DAG can be generated by activation of phospholipase D (PLD), which cleaves phosphatidylcholine to produce phosphatidic acid and choline. Dephosphorylation of phosphatidic acid produces DAG. The PLD pathway may be used by some mitogens and growth factors and likely contains a variety of activation schemes that may include G proteins.

ADDITIONAL LIPID MESSENGERS

DAG is itself a source of another lipid messenger, by the action of phospholipase A_2 (PLA_2), which releases

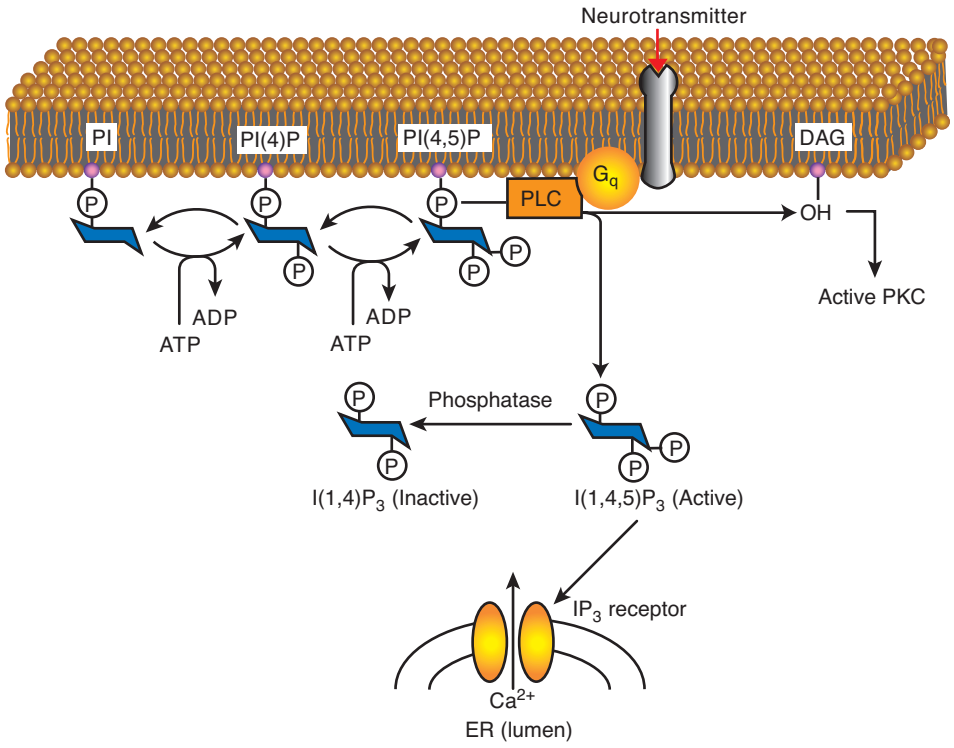


FIGURE 9.7 Schematic pathway of IP3 and DAG synthesis and action. Stimulation of receptors coupled to Gq activates PLCβ, which leads to the release of DAG and IP_3. DAG activates PKC, whereas IP_3 stimulates the IP_3 receptor in the endoplasmic reticulum (ER), leading to mobilization of intracellular Ca^{2+} stores.

the fatty acid, typically arachidonic acid, from the *sn*-2 position of the DAG backbone (Fig. 9.6). Arachidonic acid has biological activity of its own in addition to serving as a precursor for prostaglandins and leukotrienes. Arachidonic acid and other *cis*-unsaturated fatty acids can modulate K^+ channels, PLCγ, and some forms of PKC.

The phosphatidylinositides exist with modifications at the 3, 4, and 5 position to generate a family of seven lipids, [PI3P, PI4P, PI5P, PI(3,4)P2, PI(4,5)P2, PI(3,5)P2, and PI(3,4,5)P3] lipids that mediate important structural or enzymatic effects. They bind to specific protein binding domains to regulate vesicular traffic and protein kinases involved in survival and cell death. Mutations in their synthesis and metabolism have pathophysiological consequences (McCrea & De Camilli, 2009). There is evidence that another lipid, sphingomyelin, is a precursor for intracellular signals as well.

IP3, A POTENT SECOND MESSENGER THAT PRODUCES ITS EFFECTS BY MOBILIZING INTRACELLULAR Ca^{2+}

The main function of IP_3 is to stimulate the release of Ca^{2+} from intracellular stores. Ca^{2+} levels are kept low in the cytosol by its sequestration in the ER where it is complexed with low-affinity-binding proteins. The ER is the major IP_3-sensitive Ca^{2+} store in cells (Fig. 9.7) and Ca^{2+} readily flows down its concentration gradient into the cell lumen upon opening of Ca^{2+} channels in the ER.

The IP_3 receptor is a macromolecular complex that functions as an IP_3 sensor and a Ca^{2+} release channel. It has a broad tissue distribution but is highly concentrated in the cerebellum. The IP_3 receptor is a tetramer of 313-kDa subunits with a single IP_3-binding site at its N-terminal of each subunit, facing the cytoplasm. Ca^{2+} release by IP_3 is highly cooperative so that a small change in IP_3 has a large effect on Ca^{2+} release from the ER. The mouse mutants *pcd* and *nervous* have deficient levels of the IP_3 receptor and exhibit defective Ca^{2+} signaling, and a genetic knockout of the IP_3 receptor leads to motor and other deficits.

TERMINATION OF THE IP_3 SIGNAL

IP_3 is a transient signal terminated by dephosphorylation to inositol. Inactivation is initiated either by dephosphorylation to inositol 1,4-bisphosphate (Fig. 9.8) or by an initial phosphorylation to a tetrakisphosphate form that is dephosphorylated by a different pathway. Both pathways have in common an enzyme that cleaves the phosphate on the 1 position. Complete dephosphorylation yields inositol, which is recycled in the biosynthetic pathway. Recycling is important because most tissues do not contain *de novo* biosynthetic pathways for making inositol. Salvaging inositol may be particularly important

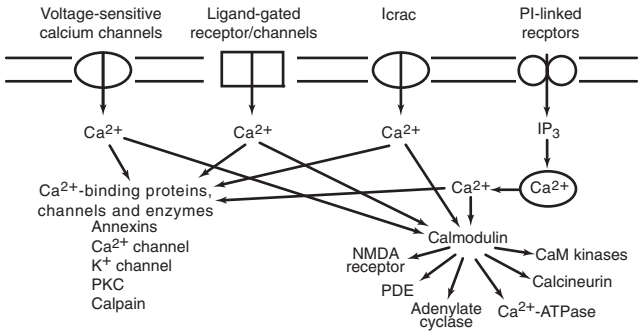


FIGURE 9.8 Multiple sources of Ca^{2+} converge on calmodulin and other Ca^{2+}-binding proteins. Cellular levels of Ca^{2+} can rise either by influx (e.g., through voltage-sensitive channels or ligand-gated channels) or by redistribution from intracellular stores triggered by IP_3. Calcium modulates dozens of cellular processes by the action of the Ca^{2+}-calmodulin complex on many enzymes, and calcium has some direct effects on enzymes such as PKC and calpain. CaM kinase, Ca^{2+}-calmodulin-dependent kinase.

when cells are actively undergoing PI turnover. It is intriguing that Li^+, the simple salt used to treat bipolar disorder, selectively inhibits the salvage of inositol by inhibiting the enzyme that dephosphorylates the 1 position and is common to the two pathways. At therapeutic doses of Li^+, the reduced salvage of inositol in cells with high phosphoinositide signaling may lead to a depletion of PIP_2 and a selective inhibition of this signaling pathway in active cells.

Calcium Ion

Calcium has a dual role as a carrier of electrical current and as a second messenger. Its effects are more diverse than those of other second messengers such as cAMP and DAG because its actions are mediated by a much larger array of proteins, including protein kinases (Carafoli and Klee, 1999). Furthermore, many signaling pathways directly or indirectly increase cytosolic Ca^{2+} concentration from 100 nM to 0.5–1.0 mM. The source of elevated Ca^{2+} can be either the ER or the extracellular space (Fig. 9.8). In addition to IP_3-mediated release, Ca^{2+} can activate its own mobilization through the ryanodine receptor on the ER. Mechanisms for Ca^{2+} influx from outside the cell include several voltage-sensitive Ca^{2+} channels and ligand-gated cation channels that are permeable to Ca^{2+} [e.g., nicotinic receptor and *N*-methyl-d-aspartate (NMDA) receptor].

DYNAMICS OF Ca^{2+} SIGNALING REVEALED BY FLUORESCENT Ca^{2+} INDICATORS

We know a great deal about the spatial and temporal regulation of Ca^{2+} signals because of the development of fluorescent Ca^{2+} indicators. A variety of fluorescent compounds selectively bind Ca^{2+} at physiological concentration ranges and rapidly change their fluorescent properties upon binding Ca^{2+} to a fairly accurate measurement of ionized Ca^{2+}.

Calmodulin-Mediated Effects of Ca^{2+}

Ca^{2+} acts as a second messenger to modulate the activity of many mediators. The predominant mediator of Ca^{2+} action is calmodulin, a ubiquitous 17-kDa calcium-binding protein. Ca^{2+} binds to calmodulin in the physiological range and converts it into an activator of many cellular targets (Cohen & Klee, 1988). Binding of Ca^{2+} to calmodulin produces a conformational change that greatly increases its affinity for more than two dozen eukaryotic enzymes that it activates, including cyclic nucleotide PDEs, adenylate cyclase, nitric oxide synthase, Ca^{2+}-ATPase, calcineurin (a phophoprotein phosphatase), and several protein kinases (Fig. 9.8). This activation of calmodulin allows neurotransmitters that change Ca^{2+} to affect dozens of cellular proteins, presumably in an orchestrated fashion.

REGULATION OF GUANYLATE CYCLASE BY NITRIC OXIDE

An important target of Ca^{2+}-calmodulin is the enzyme nitric oxide synthase (NOS). This enzyme synthesizes one of the simplest known messengers, the gas NO (Sen & Snyder, 2010). In the pathway that led to its discovery, acetylcholine stimulates the PI signaling pathway in the endothelium to increase intracellular Ca^{2+}, which activates NOS so that more NO is made. NO then diffuses radially from the endothelial cells across two cell membranes to the smooth muscle cell, where it activates guanylate cyclase to make cGMP. This in turn activates a cGMP-dependent protein kinase that phosphorylates proteins, leading to a relaxation of muscle. In 1998, Robert F. Furchgott, Louis J. Ignarro, and Ferid Murad received the Nobel Prize for their discoveries concerning nitric oxide as a signaling molecule and therapeutic mediator in the cardiovascular system.

Let us now turn to the details of the NO pathway. Nitric oxide is derived from L-arginine in a reaction catalyzed by NOS, a complex enzyme that converts L-arginine and O_2 into NO and L-citrulline. NO lasts only a few seconds in biological fluids and thus no specialized processes are needed to inactivate this particular signaling molecule. As a gas, NO is soluble in both aqueous and lipid media and can diffuse readily from its site of synthesis across the cytosol or cell membrane and affect targets in the same cell or in nearby neurons, glia, and vasculature (Sen & Snyder, 2010). NO produces a variety of effects, including relaxation of smooth muscle of the vasculature, relaxation of smooth muscle of the gut in peristalsis, and killing of foreign cells by macrophages. It was first recognized as a neuronal messenger that couples glutamate receptor stimulation to increases in cGMP. NO produced by Ca^{2+}-calmodulin-dependent

activation of NOS concentrated in cerebellar granule cells activates guanylate cyclase in nearby Purkinje cells during the induction of long-term depression in the cerebellum (see Chapter 32).

ACTIVATION OF GUANYLATE CYCLASES

Two types of guanylate cyclase, a soluble one regulated by NO and a membrane-bound enzyme regulated directly by neuropeptides, synthesize cGMP from GTP in a reaction similar to the synthesis of cAMP from ATP. NO activates the soluble enzyme by binding to the iron atom of the heme moiety. This is the basic mechanism for the regulation of soluble guanylate cyclases. A number of therapeutic muscle relaxants, such as nitroglycerin and nitroprusside, are NO donors that produce their effects by stimulating cGMP synthesis. Membrane-bound guanylate cyclases are transmembrane proteins with a binding site for neuroendocrine peptides on the extracellular side of the plasma membrane and a catalytic domain on the cytosolic side.

Cyclic GMP Phosphodiesterase, an Effector Enzyme in Vertebrate Vision

The versatility of G-protein signaling is illustrated in vertebrate phototransduction, in which a specialized G protein called transducin (G_t) is activated by light rather than by a hormone or neurotransmitter. Transducin stimulates cGMP phosphodiesterase, an effector enzyme that hydrolyzes cGMP and ultimately turns off the dark current (see Chapter 27). Nature has evolved an elegant mechanism for using photons of light to modify a hormone-like molecule, retinal, that activates a seven-helix receptor called rhodopsin. Activated rhodopsin dissociates α_t from transducin which then activates a soluble cGMP phosphodiesterase.

Rods can detect a single photon of light because the signal-to-noise ratio of the system is very low and the amplification factor in phototransduction is quite high; one rhodopsin molecule stimulated by a single photon can activate 500 transducins. Transducin remains in the "on" state long enough to activate 500 PDEs. PDE can hydrolyze about 100 cGMP molecules in the second before it is deactivated. cGMP in rods regulates a cGMP-gated cation channel, leading to additional amplification of the signal.

Modulation of Ion Channels by G Protein

Each type of neuron has a repertoire of ion channels that give it a distinct response signature, and it is not surprising that several types of mechanisms regulate these channels. Channel modulation occurs via G proteins, second messengers, and their cognate protein kinases that phosphorylate ion channels as well as by direct effects of G proteins (Greengard, 2001).

The first ion channel demonstrated to undergo regulation by G proteins was the cardiac K^+ channel that mediates slowing of the heart by acetylcholine released from the vagus nerve. When this I_{KACh} channel is examined in a membrane patch delimited by the seal of a cell-attached electrode, the addition of acetylcholine within the electrode increases the frequency of channel opening dramatically, whereas the addition of acetylcholine to the cell surface outside the seal does not. The process is therefore described as membrane delimited, with a direct interaction between the G protein (either α_i or $\beta\gamma$) and the channel.

Ca^{2+} channels are also modulated by G proteins. The central role played by Ca^{2+} in muscle contraction, in synaptic release, and in gene expression makes Ca^{2+} influx a compelling target for regulation by neurotransmitters. In the heart, where L-type Ca^{2+} channels are critical for the regulation of contractile strength, the Ca^{2+} current is enhanced by a_s formed by β-adrenergic stimulation of G_s. In contrast, N-type Ca^{2+} channels, which modulate synaptic release in nerve terminals, are often inhibited by muscarinic and α-adrenergic agents and by opiates acting at receptors coupled to G_i and G_o.

G-Protein Signaling Gives Special Advantages in Neural Transmission

The G-protein-based signaling system provides several advantages over fast transmission (Hille, 1992; Birnbaumer, 2007). These advantages include amplification of the signal, modulation of cell function over a broad temporal range, diffusion of the signal to a large cellular volume, cross talk, and coordination of diverse cell functions. The sacrifice in speed relative to signaling by ligand gated ion channels is compensated by a broad range of signaling that facilitates integration of signals by the G-protein system. A slower time frame means that cellular processes that are quite distant from the receptor can be modulated. Diffusion of second messengers such as IP_3, Ca^{2+}, and DAG can extend neurotransmission through the cell body and to the nucleus to alter gene expression and via NO to other cells. Neurotransmitters acting through G proteins can elicit a coordinated response of the cell that can modulate synaptic release, resynthesis of neurotransmitter, membrane excitability, the cytoskeleton, metabolism, and gene expression.

Summary

A major class of signaling utilizing G-protein-linked signals affords the nervous system a rich diversity of modulation, amplification, and plasticity. Signals are mediated through second messengers activating proteins that modify cellular processes and gene transcription. A key feature is the ability of G proteins to detect the presence of activated receptors and to amplify the signal

through effector enzymes and channels. Phosphorylation of key intracellular proteins, ion channels, and enzymes activates diverse, highly regulated cellular processes. The specificity of response is ensured through receptors reacting only with a limited number of G proteins. Coupling between receptor, G protein, and its effector(s), and spatial compartmentalization of the system enables specificity and localized control of signaling. Phospholipids and phosphoinositols provide substrates for second-messenger signaling for G proteins. Stimulation of release of intracellular calcium is often the mediator of the signal. Calcium itself has a dual role as a carrier of electrical current and as a second messenger. Calmodulin is a key regulator that provides complexity and enhances specificity of the signaling system. Sensitivity of the system is imparted by an extremely robust amplification system, as seen in the visual system, which can detect single photons of light.

MODULATION OF NEURONAL FUNCTION BY PROTEIN KINASES AND PHOSPHATASES

Protein phosphorylation and dephosphorylation are key processes that regulate cellular function. They play a fundamental role in mediating signal transduction initiated by neurotransmitters, neuropeptides, growth factors, hormones, and other signaling molecules (Fig. 9.9). The functional state of many proteins is modified by phosphorylation-dephosphorylation, the most ubiquitous post-translational modification in eukaryotes. More than a fifth of all proteins serve as targets for kinases and phosphatases. Phosphorylation or dephosphorylation can rapidly modify the function of enzymes, structural and regulatory proteins, receptors, and ion channels taking part in diverse processes, without a need to change the level of their expression. It can also produce long-term alterations in cellular properties by modulating transcription and translation and changing the complement of proteins expressed by cells.

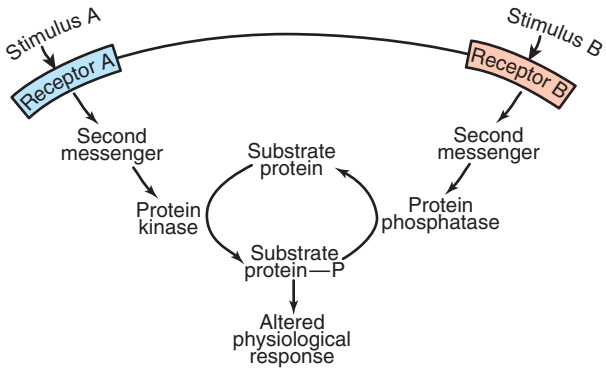


FIGURE 9.9 Regulation by protein kinases and protein phosphatases. Enzymes and other proteins serve as substrates for protein kinases and phosphoprotein phosphatases, which modify their activity and control them in a dynamic fashion. Multiple signals can be integrated at this level of protein modification. *Adapted from Svenningsson et al. (2004).*

Protein kinases catalyze the transfer of the terminal, or γ, phosphate of ATP to the hydroxyl moieties of Ser, Thr, or Tyr residues at specific sites on target proteins. Most protein kinases are either Ser/Thr kinases or Tyr kinases, with only a few designed to phosphorylate both categories of acceptor amino acids. Protein phosphatases catalyze the hydrolysis of the phosphoryl groups from phosphoserine, phosphothreonine, phosphotyrosine, or both types of phosphorylated amino acids on phosphoproteins.

The activity of protein kinases and protein phosphatases is often regulated either by a second messenger (e.g., cAMP or Ca^{2+}) or by an extracellular ligand (e.g., nerve growth factor). In general, second-messenger-regulated kinases modify Ser and Thr, whereas receptor-linked kinases modify Tyr. Among the many protein kinases and protein phosphatases in neurons, a relatively small number serve as master regulators to orchestrate neuronal function. The cAMP-dependent protein kinase (PKA) is a prototype for the known regulated Ser/Thr kinases; they are similar in overall structure and regulatory design. PKA is the predominant mediator for signaling through cAMP, the only other being a cAMP-liganded ion channel in olfaction and an exchange protein directly activated by cAMP (Epac), which modulate GDP–GTP exchange for the small GTPases Rap1 and Rap2. Epacs are activated by cAMP independent of PKA action and regulate cell adhesion, cell junction formation, and exocytosis.

In a similar fashion, the related cGMP-dependent protein kinase (PKG) mediates most of the actions of cGMP. Ca^{2+}-calmodulin-dependent protein kinase II (CaMKII) and several other kinases mediate many of the actions of stimuli that elevate intracellular Ca^{2+}. Finally, the PI signaling system increases both DAG and Ca^{2+}, which activate any of a family of protein kinases collectively called protein kinase C (PKC).

The activities of these second messenger-regulated protein kinases are countered by a relatively small number of phosphatases, exemplified by protein phosphatase 1 (PP–1), protein phosphatase 2A (PP–2A), and protein phosphatase 2B (PP–2B, or calcineurin). Phosphorylation and dephosphorylation are reversible processes, and the net activity of the two processes determines the phosphorylation state of each substrate. The Nobel Prize for Physiology and Medicine was awarded to Edwin G. Krebs and Edmund H. Fischer in 1992 for their pioneering work on the regulation of cell function by protein kinases and phosphatases.

Certain Principles are Common in Protein Phosphorylation and Dephosphorylation

Protein kinases and protein phosphatases are described either as multifunctional if they have a broad specificity and therefore modify many protein targets or as dedicated if they have a very narrow substrate specificity. Spatial positioning of kinases and their substrates in the cell further increases or decreases the likelihood of phosphorylation-dephosphorylation of a given substrate.

The amplification of signal transduction described earlier is continued during transmission of the signal by protein kinases and protein phosphatases. In some cases, the kinases are themselves subject to activation by phosphorylation in a cascade in which one activated kinase phosphorylates and activates a second, and so on, to provide amplification and a switch-like response termed ultrasensitivity.

Kinases and phosphatases integrate cellular stimuli and encode the stimuli as the steady-state level of phosphorylation of a large complement of proteins in the cell (Hunter, 1995). Distinct signal transduction pathways can converge on the same or different target substrates. In some cases, these substrates can be phosphorylated by several kinases at distinct sites. Phosphorylation can alter cellular processes over broad time scales, from milliseconds to hours and even much longer by altering gene expression.

Phosphorylation produces specific changes in the function of a target protein, such as increasing or decreasing the catalytic activity of an enzyme, conductance of an ion channel, or desensitization of a receptor. Kinases and phosphatases modulate proteins by regulating the presence of a highly charged and bulky phosphoryl moiety on Ser, Thr, or Tyr at a precise location on the substrate protein. The phosphate may elicit a conformational change or alter interaction with other proteins.

Finally, each of the three kinases exemplified here can function as a cognitive kinase—that is, a kinase capable of a molecular memory. Although the activity of each kinase requires a second messenger for initial activation, each can become persistently active and independent of its second messenger. This molecular memory potentiates the activity of these kinases and may enable them to participate in aspects of neuronal plasticity.

cAMP-Dependent Protein Kinase was the First Well-Characterized Kinase

Neurotransmitters that stimulate the synthesis of cAMP exert their intracellular effects primarily by activating PKA. The functions (and substrates) regulated by PKA include gene expression [cAMP response element-binding protein (CREB)], catecholamine synthesis (tyrosine hydroxylase), carbohydrate metabolism (phosphorylase kinase), cell morphology [microtubule-associated protein 2 (MAP–2)], postsynaptic sensitivity (AMPA receptor), and membrane conductance (K channel). Paul Greengard and Eric Kandel received the Nobel Prize for Medicine in 2000 (along with Arvid Carlsson) for their discoveries concerning signal transduction via PKA and phosphoprotein phosphatases in the nervous system. PKA is a tetrameric protein composed of two types of subunits: (1) a dimer of regulatory (R) subunits (either two RI subunits for type I PKA or two RII subunits for type II PKA) and (2) two catalytic subunits (C subunit). Two or more isoforms of the RI, RII, and C subunits have distinct tissue and developmental patterns of expression but appear to function similarly. The C subunits are 40-kDa proteins that contain the binding sites for protein substrates and ATP. The R subunits are 49-to 51-kDa proteins that contain two cAMP-binding sites. In addition, the R subunit dimer contains a region that interacts with cellular anchoring proteins that serve to localize PKA appropriately within the cell.

The binding of second messengers by PKA and the other second-messenger-regulated kinases relieves an inhibitory constraint and thus activates the enzymes (Fig. 9.10). cAMP binding leads to subunit dissociation, thereby relieving the C subunit of its inhibitory R subunits and activating the kinase. The steady-state level of cAMP determines the fraction of PKA that is in the dissociated or active form. In this way PKA decodes cAMP signals into the phosphorylation of proteins and the resultant change in various cellular processes.

PKA is a member of a large family of protein kinases that have in common a significant degree of homology in their catalytic domains and are likely derived from an ancestral gene (Fig. 9.11). This homology extends to the three-dimensional crystal structure based on x-ray

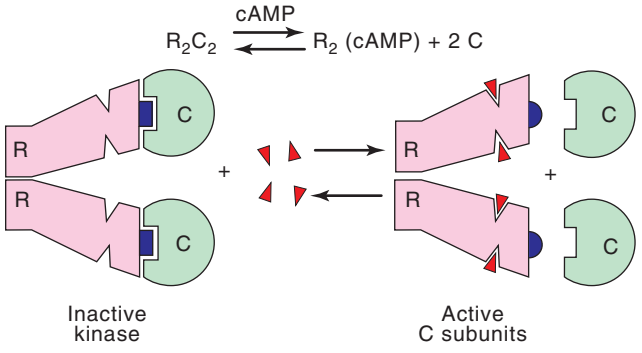


FIGURE 9.10 Activation of PKA by cAMP. An autoinhibitory segment (blue) of the regulatory subunit (R) dimer interacts with the substrate-binding domain of the catalytic (C) subunits of PKA, blocking access of substrates to their binding site. Binding of four molecules of cAMP reduces the affinity of R for C, resulting in dissociation of constitutively active C subunits.

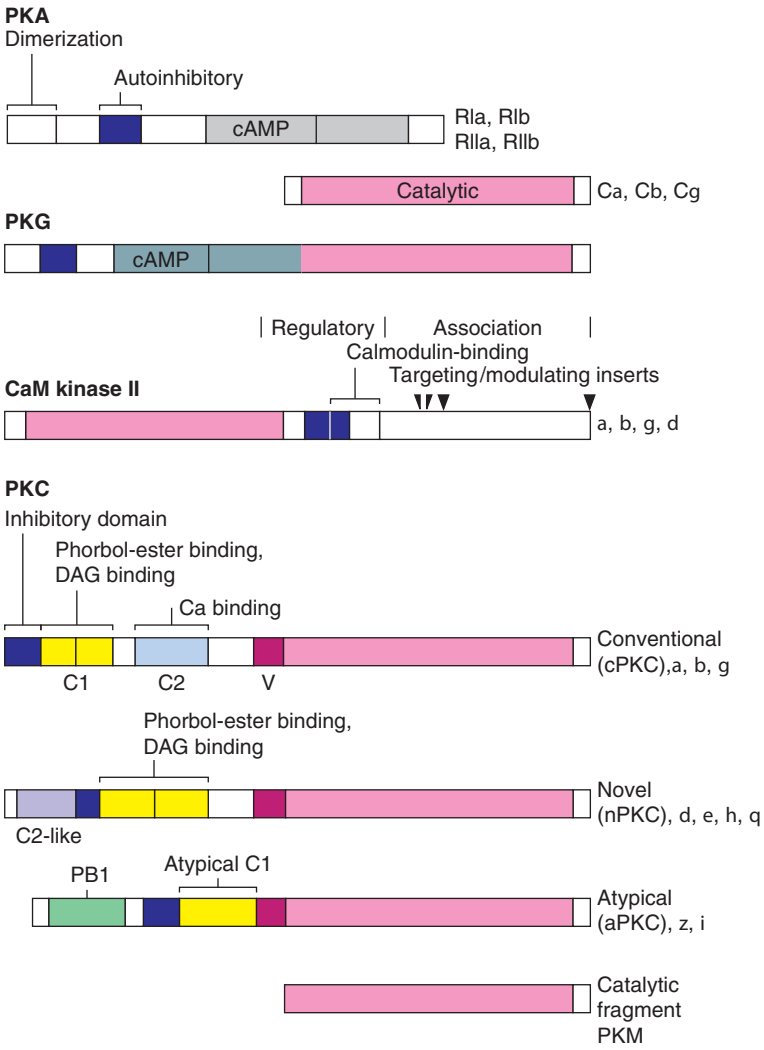


FIGURE 9.11 Domain structure of protein kinases. Protein kinases are encoded by proteins with recognizable structural sequences that encode specialized functional domains. Each of the kinases [PKA, PKG, CaMKII, and PKC] has homologous catalytic domains that are kept inactive by the presence of an autoinhibitory segment (blue lines). Regulatory domains contain sites for binding second messengers such as cAMP, cGMP, Ca^{2+}-calmodulin, DAG, and Ca^{2+}-phosphatidylserine. Alternative splicing creates additional diversity.

crystallography of PKA and other kinases. The catalytic domain may be in a subunit distinct from the regulatory domain, as in PKA, or in the same subunit, as in PKC and the CaM kinases. The crystal structure of the C subunit complexed to a segment of protein kinase inhibitor (PKI), a selective high-affinity inhibitor of PKA, reveals that the C subunit is composed of two lobes. The N-terminal lobe contains a highly conserved region that binds Mg^{2+}-ATP in a cleft between the two lobes. A larger C-terminal lobe contains the protein-substrate recognition sites and the appropriate amino acids for catalyzing transfer of the γ-phosphoryl moiety from ATP to the substrate. Inhibition by PKI is diagnostic of PKA involvement; PKI contains an autoinhibitory sequence resembling PKA substrates and is positioned in the catalytic site like a substrate, thus blocking access for substrates.

PKA phosphorylates Ser or Thr at specific sites in dozens of proteins. The sequences of amino acids at the phosphorylation sites are not identical. Each kinase has a characteristic consensus sequence that forms the basis for distinct substrate specificities.

A regulatory theme common to PKA, CaMKII, and PKC is that their second messengers activate them by displacing an inhibitory domain from the active site. In PKA, binding of the R subunit distorts the active site of the C subunit and blocks access of substrates by positioning a pseudosubstrate or inhibitory sequence in the catalytic site. Binding of cAMP to the R subunit produces a large conformation change in R that disrupts its binding to the C subunit, thus leading to dissociation of an active C subunit. CaMKII and PKC have the catalytic domain and inhibitory sequence in the same polypeptide with binding of their respective second-messengers similarly de-inhibiting the kinase by displacement of the autoinhibitory domain (Fig. 9.11).

Multifunctional CaMKII Decodes Diverse Signals That Elevate Intracellular Ca^{2+}

Most of the effects of Ca^{2+} in neurons and other cell types are mediated by calmodulin, and many of the effects of Ca^{2+}-calmodulin are mediated by protein phosphorylation-dephosphorylation (Wayman et al., 2008). The Ca^{2+}-signaling system contains a family of Ca^{2+}-calmodulin-dependent protein kinases with broad substrate specificity, including CaM kinases I, II, and IV; of these, CaMKII is the best characterized. CaMKII phosphorylates tyrosine hydroxylase, MAP-2, synapsin I, calcium channels, Ca^{2+}-ATPase, transcription factors, and glutamate receptors and thereby regulates synthesis of catecholamines, cytoskeletal function, synaptic release in response to high-frequency stimuli, calcium currents, calcium homeostasis, gene expression, and synaptic plasticity, respectively. This kinase is found in every tissue but is particularly enriched in neurons. It is found in the cytosol, in the nucleus, in association with cytoskeletal elements, and in postsynaptic thickening termed the postsynaptic density found in asymmetric synapses. It is a large multimeric enzyme, consisting of 12 subunits derived from four homologous genes (α, β, γ, and δ) that encode different isoforms of the kinase that range from 54 to 72 kDa per subunit.

The catalytic, regulatory, and targeting domains of CaMKII are all contained within a single polypeptide (Fig. 9.11). Following the catalytic domain on the N-terminal half of each isoform is the regulatory domain, which contains an autoinhibitory domain with an overlapping calmodulin-binding sequence. The C-terminal end contains an association domain that allows 12 subunits (two rings of six catalytic domains each) to assemble into a multimer, as well as targeting sequences that direct the kinase to distinct intracellular sites (Chao et al., 2011).

Regulation of the kinase by autophosphorylation is a critical feature of CaMKII. The kinase is inactive in the basal state because an autoinhibitory segment distorts the active site and sterically blocks access to its substrates. Binding of Ca^{2+}-calmodulin to the calmodulin-binding domain of the kinase displaces the autoinhibitory domain from the catalytic site, thus activating the kinase by enabling ATP and protein substrates to bind. Displacement of this domain also exposes a binding site for anchoring proteins that the activated kinase can bind. If the kinase is activated, it can autophosphorylate Thr-286 (in α-CaMKII). Phosphorylation disables the autoinhibitory segment by preventing it from reblocking the active site after calmodulin dissociates and thereby locks the kinase in a partially active state that is independent, or autonomous, of Ca^{2+}-calmodulin and can anchor to additional targets. Autophosphorylation prolongs the active state of the kinase, a potentiation that led to its description as a cognitive kinase.

CaMKII is targeted to distinct cellular compartments. Differences between the four genes encoding CaMKII and between the two or more isoforms that are encoded by each gene by apparent alternative splicing reside primarily in a variable region at the start of the association domain (Fig. 9.11). In some isoforms, this region contains an additional sequence that targets those isoforms to the nucleus. The major neuronal isoform, α-CaMKII, is largely cytosolic but is also found associated with postsynaptic densities and synaptic vesicles. Targeting to the NMDA type glutamate receptor occurs only after calmodulin activates the kinase and exposes a binding site. The kinase also serves as a scaffold for binding of other proteins, including the recruitment of the ubiquitin proteasome to dendritic spines.

Protein Kinase C is the Principal Target of the PI Signaling System

Protein kinase C is a collective name for members of a relatively diverse family of protein kinases most closely associated with the PI-signaling system. PKC is a multifunctional Ser/Thr kinase capable of modulating many cellular processes, including exocytosis and endocytosis of neurotransmitter vesicles, neuronal plasticity, gene expression, regulation of cell growth and cell cycle, ion channels, and receptors. The role of DAG generated during PI signaling was unclear until its link to PKC was established. Many PKC isoforms also require an acidic phospholipid such as phosphatidylserine for appropriate activation. The kinase is also of interest because it is the target of a class of tumor promoters called phorbol esters. They activate PKC by simulating the action of DAG, bypassing the normal receptor-based pathway, and inappropriately stimulating cell growth.

We now understand that the PKC family of kinases is diverse in structure and regulatory properties (Newton, 2010). PKC is monomeric (78–90 kDa) with catalytic, regulatory, and targeting domains all on one polypeptide (Fig. 9.11). The conventional isoforms (or cPKC), have all of the following domains:

- an autoinhibitory or pseudosubstrate sequence
- C1, a cysteine-rich domain that binds DAG and phorbol esters
- C2, a region necessary for Ca^{2+} sensitivity and for binding to phosphatidylserine and to anchoring proteins
- V, a variable protease-sensitive hinge
- the catalytic domain.

Another class of isoforms, termed novel PKCs (nPKC), lacks a true C2 domain and is therefore not Ca^{2+} sensitive. Another class is considered atypical (aPKC) because it lacks C2 and the first of two cysteine-rich domains that

are necessary for DAG (or phorbol ester) sensitivity. This class is neither Ca^{2+} nor DAG sensitive but has a PB1 domain involved in protein interaction. Not included is a family of closely related kinases termed PKN.

Activation of PKC is best understood for the conventional isoforms. Generation of DAG resulting from stimulation of the PI-signaling pathway increases the affinity of cPKC isoforms for Ca^{2+} and phosphatidylserine. DAG, or specifically its *sn*-1,2-diacylglycerol isomer, is derived only from PI turnover, and it is the only isomer effective in activating PKC. Cell stimulation results in the translocation of cPKC from a variety of sites to the membrane or cytoskeletal elements where it interacts with PS-Ca^{2+}-DAG at the membrane. Binding of the second messengers to the regulatory domain disrupts the nearby autoinhibitory domain, leading to a reversible activation of PKC by deinhibition. Translocation is not restricted to the plasma membrane. Upon activation some PKC isoforms reversibly translocate to intracellular sites enriched with anchoring proteins, termed receptors for activated C kinase (RACK).

Prolonged activation of PKC can be produced by the addition of phorbol esters, which simulate activation by DAG but remain in the cell until they are washed out. In a matter of hours to days, such persistent activation by phorbol esters leads to a degradation of PKC. This phenomenon is sometimes used experimentally to produce a PKC-depleted cell (at least for phorbol ester-binding isoforms) and thereafter to test for a loss of putative PKC functions.

Spatial Localization Regulates Protein Kinases and Phosphatases

Protein kinases and protein phosphatases are often positioned spatially near their substrates or they translocate to their substrates upon activation to improve speed and specificity in response to neurotransmitter stimulation. For example, A Kinase Anchoring Protein 79 (AKAP79) while first identified with PKA binding, also has binding site for PKC and calcineurin (Logue & Scott, 2010). Another example of a signaling complex is the protein termed yotiao, which binds to the NMDA type glutamate receptor and serves as an anchor for both PKA and a phosphatase (PP–1).

The use of anchoring proteins has several consequences. First, rate of phosphorylation and specificity are enhanced when kinases or phosphatases are concentrated near intended substrates. Second, it increases the signal-to-noise ratio for substrates that are not near anchoring proteins by reducing basal state phosphorylation. For example, PKA is anchored on the Golgi away from the nucleus so that phosphorylation in the basal state or even after a brief stimulus produces little phosphorylation of nuclear proteins. Prolonged stimuli, however, enable some C subunits to diffuse passively through nuclear pores and regulate gene expression. Termination of the nuclear action of C subunits is aided by PKI, which acts to inhibit and export it back out of the nucleus. Third, anchoring can enable significant phosphorylation of nearby substrates at basal cAMP, such as a Ca^{2+} channel phosphorylated when its phosphorylation site is exposed during depolarization.

The Cognitive Kinases

The ability of three major Ser/Thr kinases (PKA, CaMKII, and PKC) in brain to initiate or maintain synaptic changes that underlie learning and memory may require that they themselves undergo some form of persistent change in activity. Both their functional and molecular properties led to their description as cognitive kinases.

cAMP-Dependent Protein Kinase

A role for PKA as a cognitive kinase can be seen in long-term facilitation of the gill-withdrawal reflex in *Aplysia* and in long-term potentiation in the rodent hippocampus. In motor neuron cultures, repeated or prolonged exposure to serotonin or cAMP leads to long-term facilitation because PKA becomes persistently active despite the fact that cAMP is no longer elevated (Chain et al., 1999). During such activation there is a preferential degradation and decrease in the inhibitory RII subunits and thus a slight excess of C subunits that remain persistently active because of insufficient RII subunits. The C subunit then enters the nucleus and induces expression of one protein that facilitates further proteolysis of RII. In this interesting process, a molecular memory of appropriate stimulation by serotonin is encoded by a persistence of PKA activity that is regenerative.

Ca^{2+}-Calmodulin-Dependent Protein Kinase

CaMKII has features of a cognitive kinase because it has a molecular memory based on autophosphorylation and it phosphorylates proteins that modulate synaptic plasticity (Lisman, Schulman, & Cline, 2002). The biochemical properties of CaMKII suggest mechanisms by which appropriate stimulus frequencies can generate an autonomous enzyme (Fig. 9.12). At low stimulus frequency, the time between stimuli is sufficient for calmodulin to dissociate and the kinase to be dephosphorylated, and the same submaximal activation will occur with each stimulus. At higher frequencies, however, some subunits will remain autophosphorylated and bound to calmodulin so successive stimuli will result in more calmodulin bound per holoenzyme, which will make autophosphorylation more probable because it requires two active proximate neighboring subunits. The enzyme is therefore able to decode

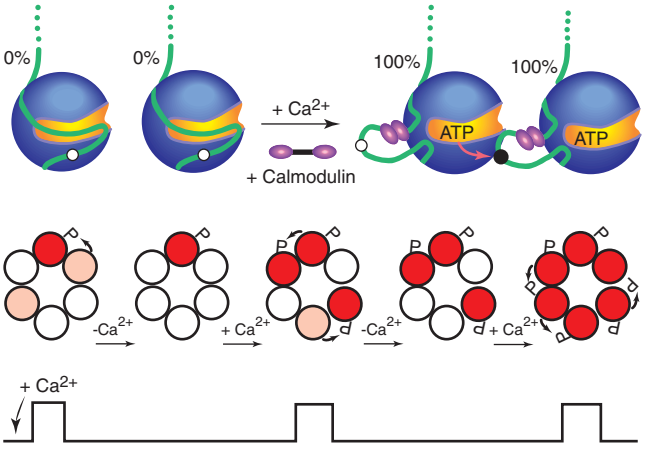


FIGURE 9.12 Frequency-dependent activation of CaMKII. Autophosphorylation occurs when both neighboring subunits in a holoenzyme are bound to calmodulin. At a high frequency of stimulation (rapid Ca^{2+} spikes), the interspike interval is too short to allow significant dephosphorylation or dissociation of calmodulin, thereby increasing the probability of autophosphorylation with each successive spike. In a simplified CaMKII with only one ring of six subunits, calmodulin-bound subunits are shown in pink and autophosphorylated subunits with trapped calmodulin are shown in red. *Adapted from Hudmon and Schulman (2002).*

the frequency of cellular stimulation and translate this into a prolonged activated state.

CaMKII phosphorylates a number of substrates that affect synaptic strength. Inhibition of CaMKII in hippocampal slices or just elimination of its autophosphorylation by an α-CaMKII mouse knock-in in which the critical Thr was replaced by Ala blocks autonomy and the induction of long-term potentiation. These mice are deficient in learning spatial navigational cues, one of the functions of the rodent hippocampus. The basis for its role involves a net shift of AMPA receptors to the synapse that may be mediated by phosphorylation of both GluA1 receptors and their accessory protein(s), leading to a greater postsynaptic response (Lisman, Schulman, & Cline, 2002; Opazo et al., 2010). The autophosphorylated kinase also serves as a scaffold to recruit the ubiquitin proteasome to synaptic sites that enable dynamic changes at the synapse.

Protein Kinase C

PKC can also be converted into a form that is independent, or autonomous, of its second messenger and can be described as a cognitive kinase. During the persistent phase of long-term potentiation, some of the inhibitory domain of some PKC molecules is proteolytically removed, thus converting it to a constitutively active kinase termed protein kinase M (PKM). PKCζ can also be transcribed from an internal promoter to produce PKM, whose translation in dendrites is unblocked during LTP and whose activity then disables the translational block, an interesting form of molecular memory (Sacktor, 2011). PKC (and PKM) substrates associated with long-term potentiation include NMDA and AMPA receptors.

Protein Tyrosine Kinases Take Part in Cell Growth and Differentiation

Protein kinases that phosphorylate tyrosine residues are usually associated with the regulation of cell growth and differentiation. Signal transduction by protein tyrosine kinases often includes a cascade of kinases phosphorylating other kinases, eventually activating Ser/Thr kinases, which carry out the intended modification of a cellular process. There are both receptor tyrosine kinases, activated by the binding of extracellular growth factors such as nerve growth factor and epidermal growth factor, and soluble ones, activated indirectly by extracellular ligands such as c-Src (Lemmon & Schlessinger, 2010).

Protein Phosphatases Undo What Kinases Create

Protein phosphatases in neuronal signaling are categorized as either phosphoserine-phosphothreonine phosphatases (PSPs) or phosphotyrosine phosphatases (PTPs) (Hunter, 1995; Mansuy & Shenolikar, 2006). The enzymes catalyze the hydrolysis of the ester bond of the phosphorylated amino acids to release inorganic phosphate and the unphosphorylated protein. A limited number of multifunctional PSPs account for most of such phosphatase activity in cells. They are categorized based on their domain structure, inhibitor sensitivity, and catalytic mechanism (Table 9.1). There are three families of protein serine/threonine phosphatases, designated as phosphoprotein phosphatases (PP-1, -2B (calcineurin), -2A, -4, -6, -5, -7), metal-dependent phosphatases (PP2C), and aspartate-based phosphatases (FCP and SCP). Protein phosphatase 2B (PP-2B, or calcineurin) responds directly to a second messenger, Ca^{2+}. The specificity of PP-1 and PP-2A is particularly broad in isolation, but both their protein substrates and cellular localization are refined by interaction with many regulatory and scaffold proteins. Phosphotyrosine phosphatases constitute a distinct and larger class of phosphatases, including PTPs with dual specificity for both phosphotyrosines and phosphoserine-phosphothreonines. PTPs are either soluble enzymes or membrane proteins with variable extracellular domains that enable regulation by extracellular binding of either soluble or membrane-bound signals.

Structure and Regulation of PP-1 and Calcineurin

PP-1 and calcineurin are the best characterized phosphatases with regard to both structure and regulation.

TABLE 9.1 Categories of Protein Phosphatases[a]

Phosphatase	Characteristic	Other Inhibitors
PP-1	Sensitive to phospho-inhibitor-1, phospho-DARPP-32, and inhibitor-2	Weakly sensitive to okadaic acid
PP-2B (calcineurin)	Ca^{2+}/calmodulin-dependent	FK506, cyclosporin
PP-2A	Highly abundant	Highly sensitive to okadaic acid
PP-4/PP-6	Nuclear	Highly sensitive to okadaic acid
PP-5	Nuclear	Mildly sensitive to okadaic acid
PP-7	Nuclear	Mildly sensitive to okadaic acid
PP-2C	Requires metal (Mg^{2+}/Mn^{2+})	EDTA
FCP/SCP[b]	Aspartic acid-based catalysis	
Receptor PTPs[c]	Plasma membrane	Vanadate
Nonreceptor PTPs	Various cellular compartments	Vanadate
Dual specificity PTPs	Nuclear (e.g., cdc25A/B/C and VH family)	Vanadate

[a]*Updated from Hunter (1995).*
[b]*TFIIF-associated component of RNA polymerase II CTD phosphatase/small CTD phosphatase.*
[c]*Protein phosphotyrosine phosphatases.*

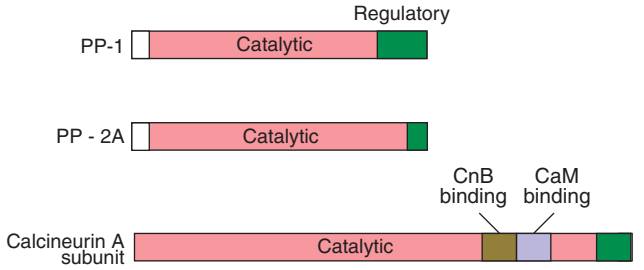


FIGURE 9.13 Domain structure of the catalytic subunits of some Ser/Thr phosphatases. The three major phosphoprotein phosphatases, PP-1, PP-2A, and calcineurin, have homologous catalytic domains but differ in their regulatory properties.

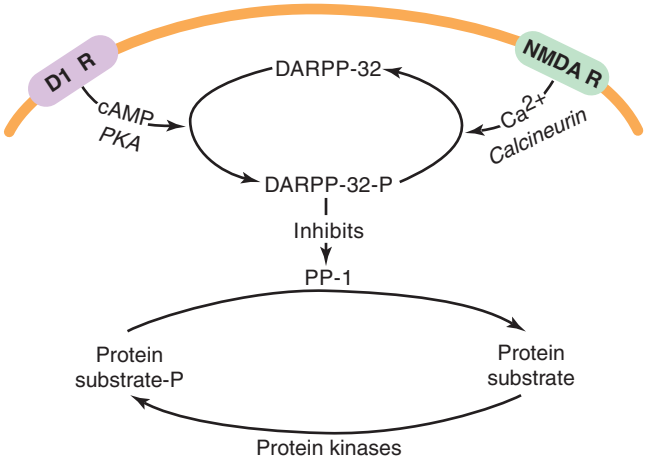


FIGURE 9.14 Cross talk between kinases and phosphatases. The state of phosphorylation of protein substrates is regulated dynamically by protein kinases and phosphatases. In the striatum, for example, dopamine stimulates PKA, which converts DARPP-32 into an effective inhibitor of PP-1. This increases the steady-state level of phosphorylation of a hypothetical substrate subject to phosphorylation by a variety of protein kinases. This action can be countered by NMDA receptor stimulation by another stimulus that increases intracellular Ca^{2+} and activates calcineurin. PP-1 is deinhibited and dephosphorylates the phosphorylated substrate when calcineurin deactives DARPP-32-P. *Adapted from Svenningsson et al. (2004).*

The domain structures of the catalytic subunits of PP-1 and calcineurin are depicted in Figure 9.13. PP-1 is a protein of 35–38 kDa; most of the sequence forms the catalytic domain, and its C-terminal is the site of regulatory phosphorylation. The catalytic domains of PP-1, PP-2A, and calcineurin are highly homologous.

PP-1 and PP-2A are normally complexed in cells with specific targeting or regulatory subunits. Targeting of PP-1 can be modulated by phosphorylation of its targets. As PP-1 dissociates from targeting subunits, it becomes susceptible to inhibition by inhibitor-2.

Inhibition of PP-1 by two other inhibitors, inhibitor-1 and its homologue DARPP-32 (dopamine and cAMP-regulated phosphoprotein; M_r 32,000), is conditional on their phosphorylation by either PKA or PKG (Fig. 9.14). Because the substrates for PKA and PP-1 overlap to a great extent, the rate and extent of phosphorylation of such substrates are enhanced by the ability of PKA to catalyze their phosphorylation while blocking their dephosphorylation via PP-1. Inhibitor-1, DARPP-32, and inhibitor-2 are all selective for PP-1. Highly selective inhibitors capable of penetrating the cell membrane are available for these phosphatases. Okadaic acid, a natural product of marine dinoflagellates, is a tumor promoter but, unlike phorbol esters, it acts on PP-2A and PP-1 rather than on PKC.

PROTEIN PHOSPHATASE 1

The x-ray structure of the catalytic subunit of PP-1 bound to the toxin microcystin, a cyclic peptide inhibitor,

reveals PP-1 to be a compact ellipsoid with hydrophobic and acidic surfaces forming a cleft for binding substrates. PP-1 is a metalloenzyme requiring two metals in the active site that likely take part in electrostatic interactions with the phosphate on substrates that aid in catalyzing the hydrolytic reaction. Substrate binding is blocked when phospho-inhibitor-1 or microcrystin LR binds to this surface.

CALCINEURIN (PP-2B)

Calcineurin is a Ca^{2+}-calmodulin-dependent phosphatase that is highly enriched in the brain. It is a heterodimer with a 60-kDa subunit (CnA) that contains an N-terminal catalytic domain similar to PP-1 and a C-terminal regulatory domain that includes an autoinhibitory segment, a calmodulin-binding domain, and a binding site for the 19-kDa regulatory B subunit (CnB). CnB is a calmodulin-like Ca^{2+}-binding protein that binds to a hinge region of CnA. Some activation of calcineurin is attained by binding of Ca^{2+} to CnB. Stronger activation is obtained by the binding of Ca^{2+}-camodulin. Additional regulation may be accorded by interaction of its hinge region with cyclophilin and FKBP (FK506-binding protein), proteins that bind the immunosuppressive agents cyclosporin and FK506, respectively.

The Ca^{2+}-calmodulin sensitivities of calcineurin and CaMKII are quite different. Weak or low-frequency stimuli may selectively activate calcineurin, whereas strong or high-frequency stimuli activate CaMKII and calcineurin. This difference may play a role in the bidirectional control of synaptic strength (depression vs. potentiation) by low- and high-frequency stimulation.

Protein Kinases, Protein Phosphatases, and Their Substrates are Integrated Networks

Cross talk between protein kinases and protein phosphatases is critical to their ability to integrate inputs into neurons. Such cross talk is exemplified by the interaction of cAMP and Ca^{2+} signals through PKA and calcineurin, respectively. The medium spiny neurons in the neostriatum receive cortical inputs from glutamatergic neurons that are excitatory and nigral inputs by dopaminergic neurons that inhibit them. A possible signal transduction scheme for this regulation is shown in Figure 9.14. The key to regulation is the bidirectional control of DARPP-32 phosphorylation (Svenningsson et al., 2004). Glutamate activates calcineurin by increasing intracellular Ca^{2+}, leading to the dephosphorylation and inactivation of phospho-DARPP-32. This releases inhibition of PP-1, which can then dephosphorylate a variety of substrates, including Na^+, K^+-ATPase, and lead to membrane depolarization. This is countered by dopamine, which stimulates cAMP formation and activation of PKA, which then converts DARPP-32 into its phosphorylated (i.e., PP-1 inhibitory) state. There are many other receptors and signaling integrated by these pathways. For example, adenosine, serotonin, and VIP act through their cognate receptors to elevate cAMP, similarly to dopamine at D1 receptors, while opiates can signal to inhibit the action of dopamine at D1 and adenosine at A2A receptors. Although PKA and calcineurin are acting in an antagonistic manner, they are not doing it by phosphorylating and dephosphorylating the ATPase. By their actions upstream, at the level of DARPP-32, the regulation of numerous target enzymes (e.g., Ca^{2+} channels and Na^+ channels) in addition to the ATPase can be coordinated.

Studying Cellular Processes Controlled by Phosphorylation-Dephosphorylation

Major goals of signal transduction research are to delineate pathways by which signals such as neurotransmitters transduce their signals to modify cellular processes. This is often the start of a process to identify targets for therapeutic intervention in disease. Cellular and biochemical assays can often identify the entire signaling pathway, from stimulation of receptor, to generation of a second-messenger activation of a kinase or phosphatase, change in the phosphorylation state of the substrate, and an ultimate change in its functional state. Such investigations utilize a variety of pharmacological inhibitors or activators of the signaling molecules complemented by genetic approaches that utilize transfection of activated forms of the kinases or phosphatases in question, siRNAs, transgenic animals, and mice with individual signaling components knocked out.

Summary

The morphology of a cell is determined by protein constituents. Its function is regulated by the phosphorylation or dephosphorylation of the proteins. Phosphorylation modifies the function of regulatory proteins subsequent to their genetic expression. The activities of the protein kinases and protein phosphatases are typically regulated by second messengers and extracellular ligands. Kinases and phosphatases integrate and encode stimulation of a large group of cellular receptors. The number of possible effects is almost limitless and enables the tuning of cellular processes over a broad time scale. Most of the effects of Ca^{2+} in cells are mediated by calmodulin, which in turn mediates changes in protein phosphorylation-dephosphorylation. The phosphoinositol signaling system is mediated through PKC, which modulates many cellular processes

from exocytosis to gene expression. All three classes of enzymes discussed have been described as cognitive kinases because they are capable of sustaining their activated states after their second-messenger stimuli have returned to basal levels. PKA has been implicated in learning and memory in *Aplysia* and in hippocampus, where it is involved in long-term potentiation. Protein phosphatases play an equally important role in neuronal signaling by dephosphorylating proteins. Cross talk between protein kinases and protein phosphatases is key to their ability to integrate inputs into neurons.

INTRACELLULAR SIGNALING AFFECTS NUCLEAR GENE EXPRESSION

For all living cells, regulation of gene expression by intracellular signals is a fundamental mechanism of development, homeostasis, and adaptation to the environment. Protein phosphorylation and regulation of gene expression by intracellular signals are the most important mechanisms underlying the remarkable degree of plasticity exhibited by neurons. Alterations in gene expression underlie many forms of long-term changes in neural functioning, with a time course that ranges from hours to many years. Signaling can also regulate translation of mRNA, including translation that is localized to dendrites and synaptic spines, as well as protein degradation that facilitates dynamic regulation of cell function and shape.

Interactions of Specific DNA Sequences with Regulatory Proteins Control Both Basal and Signal-Regulated Transcription

Information contained within DNA must be expressed through other molecules: RNA and proteins. The human genome contains approximately 25,000 genes that encode structural RNAs or protein-coding messenger RNAs (mRNAs). Regulated gene expression conferred by the nucleotide sequence of the DNA itself is called *cis* regulation because the control regions are linked physically on the DNA to regions that can potentially be transcribed. The *cis* regulatory sequences function by serving as high-affinity binding sites for regulatory proteins called transcription factors.

The transcription of specific genes into mRNA is carried out by a complex enzyme called RNA polymerase II. Roger Kornberg won the Nobel Prize in Chemistry in 2006 for his studies on the molecular basis for eukaryotic transcription. Transcription is often divided into three steps: initiation of RNA synthesis, RNA chain elongation, and chain termination. Extracellular signals, such as neurotransmitters, hormones, drugs, and growth factors generally control the transcription initiation step.

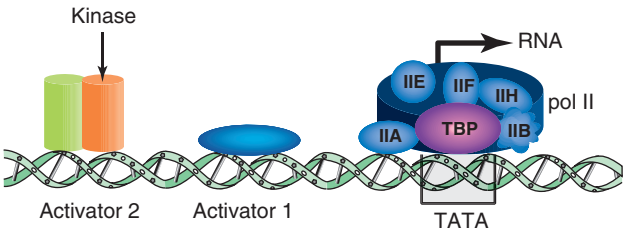


FIGURE 9.15 Schematic of a generalized RNA polymerase II promoter showing three separate *cis*-regulatory elements along a stretch of DNA. These elements are two hypothetical activator protein-binding sites and the TATA element. The TATA element is shown binding the TATA-binding protein (TBP). Multiple general transcription factors (IIA, IIB, etc.) and RNA polymerase II (pol II) associate with TBP. Each transcription factor comprises multiple individual proteins complexed together. This basal transcription apparatus recruits RNA polymerase II into the complex and also forms the substrate for interactions with the activator proteins binding to the activator elements shown. Activator 2 is shown to be a substrate for a protein kinase.

Transcription initiation requires two critical processes: (1) positioning of RNA polymerase II at the correct start site of the gene to be transcribed and (2) controlling the efficiency of initiations to produce the appropriate transcriptional rate for the circumstances of the cell (Tjian & Maniatis 1994). The *cis*-regulatory elements that set the transcription start sites of genes are called the basal promoter. Other *cis*-regulatory elements tether additional activator and represser proteins to the DNA to regulate the overall transcriptional rate (Fig. 9.15).

Sequence-Specific Transcription Factors

The promoters for RNA polymerase II transcribed genes contain a distinct basal promoter element on which a basal transcription complex is assembled. The basal promoter of most of these genes contains a sequence called a TATA box that is rich in the nucleotides adenine (A) and thymine (T) located between 25 and 30 bases upstream of the transcription start site.

To achieve significant levels of transcription, this multiprotein assembly requires help from sequence-specific transcriptional activators that recognize and bind distinct *cis*-regulatory elements. Functional *cis*-regulatory elements are generally 7–12 bp in length and structured as a palindrome, each of which is a specific binding site for one or more transcription factors. Each gene has a particular combination of *cis*-regulatory elements, the nature, number, and spatial arrangement of which determine the gene's unique pattern of expression, including the cell types in which it is expressed, the times during development in which it is expressed, and the level at which it is expressed in adults both basally and in response to physiological signals (Tjian & Maniatis, 1994).

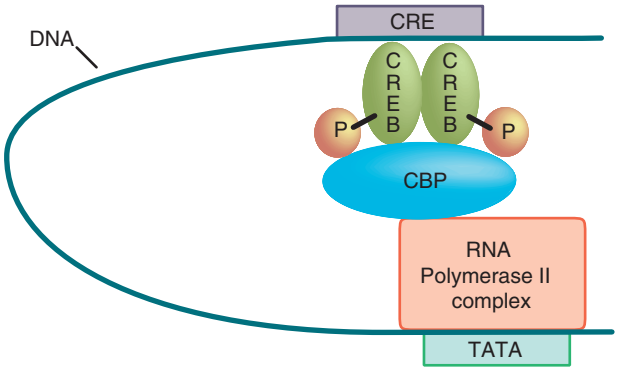


FIGURE 9.16 Looping of DNA permits activator (or represser) proteins binding at a distance to interact with the basal transcription apparatus. The basal transcription apparatus is shown as a single box (pol II complex) bound at the TATA element. The activator protein (CREB) is shown as having been phosphorylated. On phosphorylation, many activators, such as CREB, are able to recruit adaptor proteins that mediate between the activator and the basal transcription apparatus. An adaptor protein that binds phosphorylated CREB is called a CREB-binding protein (CBP).

Many transcription factors are active only as dimers or higher order complexes formed via a multimerization domain. Both partners in a dimer commonly contribute jointly to both the DNA-binding domain and the activation domain. Dimerization can be a mechanism of either positive or negative control of transcription. The effects of sequence-specific transcriptional activator and represser factors are frequently mediated by adapter proteins (Fig. 9.16). In many cases, these adapter proteins are enzymes, such as histone acyl transferases and deacetylases, with the ability to modify the structure of the proteins associated with the DNA and modulate transcription (Fig. 9.16).

A Significant Consequence of Intracellular Signaling is the Regulation of Transcription

Intracellular signals play a major role in the regulation of gene expression—for example, via nuclear translocation and/or phosphorylation of activator proteins. Signal-directed change in location or conformation of these proteins permits information obtained by the cell from its different signaling systems to regulate gene expression appropriate to the status of the cell.

Transcriptional Regulation by Intracellular Signals

Extracellular control of transcription requires a translocation step by which the signal is transmitted through the cytoplasm to the nucleus. Some transcription factors are themselves translocated to the nucleus. For example, the transcription factor NF-κB is retained in the cytoplasm by its binding protein IκB, which masks the NF-κB nuclear localization signal. Signal-regulated phosphorylation of IκB by PKC and other protein kinases leads to dissociation of NF-κB, permitting it to enter the nucleus. Other transcription factors must be directly phosphorylated or dephosphorylated to bind DNA. In many cytokine-signaling pathways, plasma membrane receptor tyrosine phosphorylation of transcription factors known as signal transducers and activators of transcription (STATs) permits their multimerization, which in turn permits both nuclear translocation and construction of an effective DNA-binding site within the multimer. Yet other transcription factors, such as CREB (Fig. 9.16), are already bound to their cognate *cis*-regulatory elements and become able to activate transcription after phosphorylation by a kinase that translocates to the nucleus.

Role of cAMP and Ca^{2+} in the Activation Pathways of Transcription

The cAMP second-messenger pathway regulates expression of a large number of genes via cAMP response elements (CREs) with the consensus palindromic sequence of TGACGTCA. Phosphorylation of CRE-bound CREB on its Ser-133 by activated PKA that translocates to the nucleus recruits the adapter protein CBP. This, in turn, interacts with the basal transcription complex and modifies histones to enhance the efficiency of transcription. CREB also serves to illustrate the convergence of signaling pathways (Fig. 9.17). CREB Ser-133 phosphorylation is not only mediated by PKA but also by CaMKII and CaMKIV and by RSK2, a kinase activated in growth factor pathways, including Ras and MAP kinase (Sakamoto, Karelina & Obrietan, 2011).

CREB illustrates yet another important principle of transcriptional regulation: CREB is a member of a family of related proteins. Many transcription factors are members of families; this permits complex forms of positive and negative regulation. CREB is closely related to other proteins called activating transcription factors (ATFs) and CRE modulators (CREMs). The dimerization domain used by CREB-ATF proteins and several other families of transcription factors is called a leucine zipper. The dimerization motif is an α helix in which every seventh residue is a leucine; based on the periodicity of α helices, the leucines line up along one face of the helix two turns apart. Dimerization juxtaposes the adjacent basic, DNA binding, regions of each of the partners. This combination of motifs is why this superfamily of proteins is referred to as basic leucine zipper proteins (bZIPs).

AP-1 Transcription Factors

Activator protein 1 (AP-1) is another family of bZIP transcription factors that play a central role in the regulation of neural gene expression by extracellular signals. The AP-1 family comprises multiple proteins that bind as heterodimers (and a few as homodimers) to the DNA sequence TGACTCA. While the AP-1 sequence

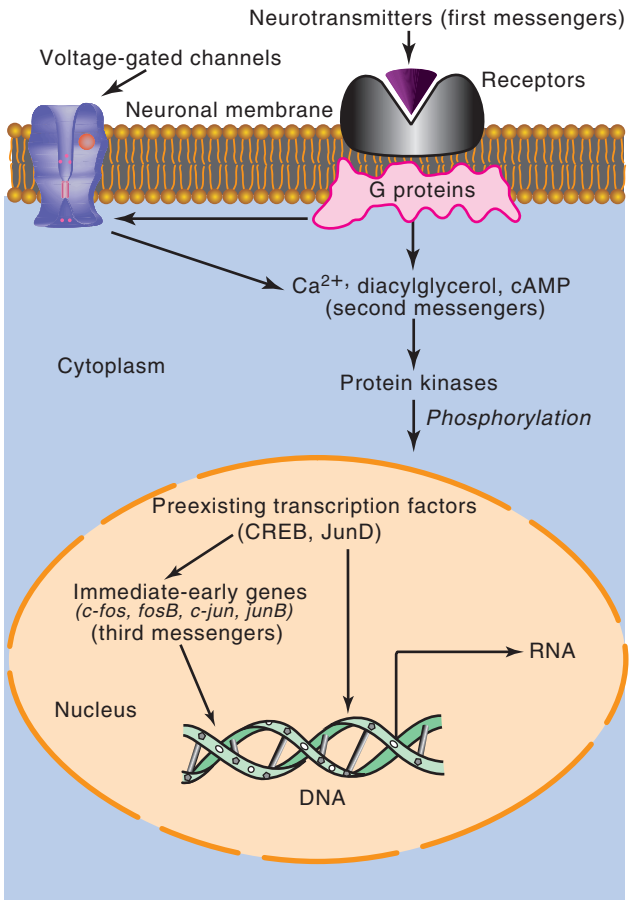


FIGURE 9.17 Signal transduction to the nucleus. In this schematic, activation of a neurotransmitter receptor activates cellular signals (G proteins and second-messenger systems). These signals, in turn, regulate the activation of protein kinases, which translocate to the nucleus. Within the nucleus, protein kinases can activate genes regulated by constitutively synthesized transcription factors. A subset of these genes encodes additional transcription factors (third messengers), which can then activate multiple downstream genes.

differs from the CRE sequence (TGACGTCA) by only a single base, this one-base difference strongly biases protein binding away from the CREB family of proteins. AP-1 sequences confer responsiveness to the PKC pathway. AP-1 proteins generally bind DNA as heterodimers composed of one member each of two different families of related bZIP proteins, the Fos family (c-Fos, Fra-1, Fra-2, and FosB) and the Jun family (c-Jun, JunB, and JunD) providing for a multiplicity of regulatory control.

Cellular Immediate-Early Genes

Genes that are activated transcriptionally by synaptic activity, drugs, and growth factors have often been classified roughly into two groups. Genes, such as the c-*fos* gene itself, that are activated rapidly (within minutes), transiently, and without requiring new protein synthesis are often described as cellular immediate-early genes (IEGs). Genes that are induced or repressed more slowly (within hours) and are dependent on new protein synthesis have been described as late-response genes. Several IEGs have been used as cellular markers of neural activation because they are markedly induced by depolarization (the critical signal being Ca^{2+} entry) and second-messenger and growth factor pathways, permitting novel approaches to functional neuroanatomy.

The protein products of those cellular IEGs that function as transcription factors bind to *cis*-regulatory elements contained within a subset of late-response genes to activate or repress them (Fig. 9.17). In sum, neural genes that are regulated by extracellular signals are activated or repressed with varying time courses by reversible phosphorylation of constitutively synthesized transcription factors and by newly synthesized transcription factors, some of which are regulated as IEGs.

*Activation of the c-*fos *Gene*

The c-*fos* gene is activated rapidly by neurotransmitters or drugs that stimulate the cAMP pathway or Ca^{2+} elevation. Both pathways produce phosphorylation of transcription factor CREB. The c-*fos* gene contains three binding sites for CREB. The c-*fos* gene can also be induced by the Ras/MAP kinase pathway, which is activated by a number of growth factors. For example, neurotrophins, such as nerve growth factor (NGF), bind a family of receptor tyrosine kinases (Trks); NGF interacts with Trk A, which activates Ras. Ras then acts through a cascade of protein kinases. Cross talk between neurotransmitter and growth factor–signaling pathways has been documented with increasing frequency and likely plays an important role in the precise tuning of neural plasticity to diverse environmental stimuli.

Regulation of c-Jun

Expression of most of the proteins of the Fos and Jun families that constitute transcription factor AP-1 and the binding of AP-1 proteins to DNA is regulated by extracellular signals. Phosphorylation and activation of c-Jun can result from the action of Jun N-terminal kinase (JNK). JNK is a member of the mitogen activated protein kinase (MAPK) family of protein kinases. JNK has also been shown to be activated by neurotransmitters, including glutamate. Thus, AP-1–mediated transcription within the nervous system requires multiple steps, beginning with the activation of genes encoding AP-1 proteins.

Cytokines as Inducers of Gene Expression in the Nervous System

With regard to function, the boundary between trophic, or growth, factors and cytokines in the nervous system has become increasingly arbitrary. However,

cell-signaling mechanisms offer a useful means of distinction. Growth factors, such as neurotrophins (e.g., nerve growth factor, brain-derived neurotrophic factor, and neurotrophin 3), epidermal growth factor (EGF), and fibroblast growth factor (FGF), act through receptor protein tyrosine kinases, whereas cytokines, such as leukemia inhibitory factor (LIF), ciliary neurotrophic factor (CNTF), and interleukin-6 (IL-6), act through nonreceptor protein tyrosine kinases.

LIF, CNTF, and IL-6 subserve a wide array of overlapping functions inside and outside the nervous system, including hematopoietic and immunologic functions outside the nervous system and regulation of neuronal survival, differentiation, and, in certain circumstances, plasticity within the nervous system. Their receptors contain a common signal-transducing subunit, gp130. Receptors for these cytokines consist of a signal-transducing β component, which includes gp130, and interacts with nonreceptor protein tyrosine kinases (PTKs) of the Janus kinase (Jak) family (e.g., Jak1, Jak2, and Tyk2). Some cytokine receptors interact only with a single Jak PTK, whereas others interact with multiple Jak PTKs.

Signal transduction to the nucleus includes tyrosine phosphorylation by the Jak PTKs of one or more of the STAT proteins mentioned earlier. Upon phosphorylation, STAT proteins form dimers through the association of SH2 domains, an important type of protein interaction domain. Dimerization is thought to trigger translocation to the nucleus, where STATs bind their cognate cytokine response elements. Different STATs become activated by different cytokine receptors, not because of differential use of Jak PTKs but because of specific coupling of certain STATs to certain receptors. Thus, for example, the IL-6 receptor preferentially activates STAT1 and STAT3; the CNTF receptor preferentially activates STAT3. Cytokine response elements, to which STATs bind, have now been identified within many neural genes, including vasoactive intestinal polypeptide and several other neuropeptide genes.

Steroid Hormone Receptors

The differentiation of many cell types in the brain is established by exposure to steroids. Steroid hormones, including glucocorticoids, sex steroids, mineralocorticoids, retinoids, thyroid hormone, and vitamin D, are small lipid-soluble ligands that can diffuse across cell membranes. They act on their receptors within the cell cytoplasm in marked distinction to the other types of intercellular signals described herein. Another unique feature of steroid hormones is that their receptors are themselves transcription factors. Each has a transcriptional-activation domain at its amino terminus, a DNA-binding domain, and a hormone-binding domain at its carboxy terminus. DNA-binding domains recognize specific palindromic DNA sequences, steroid hormone response elements, within the regulatory regions of specific genes.

After having been bound by hormone, activated steroid hormone receptors translocate into the nucleus, where they bind to their cognate response elements. Such binding then increases or decreases the rate at which these target genes are transcribed, depending on the precise nature and DNA sequence context of the element.

Summary

The formation of long-term memories requires changes in gene expression and new protein synthesis. It is at transcription initiation that extracellular signals such as neurotransmitters, hormones, drugs, and growth factors exert their most significant control. The transcription is modulated by transcription factors that recruit the RNA polymerases to the DNA. For example, the critical nuclear translocation step in the activation of transcription factor CREB involves the catalytic subunit of PKA, which can phosphorylate CREB on entering the nucleus. In addition, increasing evidence indicates that at least some forms of long-term memory require new gene expression.

Genes that encode the transcription factors themselves may respond quickly or slowly. These genes have been coined third messengers in signal transduction cascades. Cross talk between neurotransmitter and growth factor-signaling pathways is likely to play an important role in the precise tuning of neuronal plasticity to diverse environmental stimuli.

The active, mature transcription complex is a remarkable architectural assembly of RNA polymerase II, transcription factors, and adaptors assembled at the basal promoter. Cells can exert exquisite control of the genes being transcribed in a variety of situations—for example, to govern appropriate entry or exit from the cell cycle, to maintain appropriate cellular identity, and to respond appropriately to extracellular signals.

Transcription can be regulated by many different extracellular signals modulated by a large array of signaling pathways (many including reversible phosphorylation) and a complex array of typically dimeric transcription factors. In this chapter, regulation has been illustrated by only a few of the families of transcription factors. Those chosen appear to play important roles in the nervous system and illustrate many of the basic principles of gene regulation.

References

Berridge, M. J. (2009). Inositol trisphosphate and calcium signalling mechanisms. *Acta Biochimica et Biophysica, 1793*, 933–940.

Birnbaumer, L. (2007). Expansion of signal transduction by G proteins. The second 15 years or so: From 3 to 16 α subunits plus βγ dimers. *Acta Biochimica et Biophysica, 1768*, 772–793.

Bourne, H. R., & Nicoll, R. (1993). Molecular machines integrate coincident synaptic signals. *Cell, 72,* 65–75.

Carafoli, E., & Klee, C. (1999). *Calcium as a Cellular Regulator*. New York: Oxford University Press.

Chain, D. G., Casadio, A., Schacher, S., Hegde, A. N., Valbrun, M., Yamamoto, N., Goldberg, A. L., Bartsch, D., Kandel, E. R., & Schwartz, J. H. (1999). Mechanisms for generating the autonomous cAMP-dependent protein kinase required for long-term facilitation in Aplysia. *Neuron, 22,* 147–156.

Chao, L. H., Stratton, M. H., Lee, I. -H., Rosenberg, O. S., Levitz, J., Mandell, D. J., Kortemme, T., Groves, J. T., Schulman, H., & Kuriyan, J. (2011). A Mechanism for tunable autoinhibition in the structure of a human Ca^{2+}/calmodulin-dependent kinase II holoenzyme. *Cell, 146,* 732–745.

Clapham, D. E., & Neer, E. J. (1993). New roles for G-protein $\beta\gamma$-dimers in transmembrane signalling. *Nature, 365,* 403–406.

Greengard, P. (2001). The neurobiology of slow synaptic transmission. *Science, 294,* 1024–1030.

Hudmon, A., & Schulman, H. (2002). Neuronal Ca^{2+}/calmodulin-dependent protein kinase II: the role of structure and autoregulation in cellular function. *Annual Review of Biochemistry, 71,* 473–510.

Hille, B. (1992). G protein-coupled mechanisms and nervous signaling. *Neuron, 9,* 187–195.

Hunter, T. (1995). Protein kinases and phosphatases: The yin and yang of protein phosphorylation and signaling. *Cell, 80,* 225–236.

Lemmon, M. A., & Schlessinger, J. (2010). Cell signaling by receptor tyrosine kinases. *Cell, 141,* 1117–1134.

Lisman, J., Schulman, H., & Cline, H. (2002). The molecular basis of CaMKII function in synaptic plasticity and behavioural memory. *Nature Review Neuroscience, 3,* 175–190.

Logue, J. S., & Scott, J. D. (2010). Organizing signal transduction through A-kinase anchoring proteins (AKAPs). *The FEBS Journal, 277,* 4370–4375.

Mansuy, I. M., & Shenolikar, S. (2006). Protein serine/threonine phosphatases in neuronal plasticity and disorders of learning and memory. *Trends in Neuroscience, 29,* 679–686.

McCrea, H. J., & De Camilli, P. (2009). Mutations in phosphoinositide metabolizing enzymes and human disease. *Physiology (Bethesda), 24,* 8–16.

Newton, A. C. (2010). Protein kinase C: Poised to signal. *American Journal of Physiology Endocrinology and Metabolism, 298,* E395–402.

Opazo, P., Labrecque, S., Tigaret, C. M., Frouin, A., Wiseman, P. W., De Koninck, P., & Choquet, D. (2010). CaMKII triggers the diffusional trapping of surface AMPARs through phosphorylation of stargazin. *Neuron, 67,* 239–252.

Sacktor, T. C. (2011). How does PKMζ maintain long-term memory? *Nature Review Neuroscience, 12,* 9–15.

Sakamoto, K., Karelina, K., & Obrietan, K. (2011). CREB: a multifaceted regulator of neuronal plasticity and protection. *Journal of Neurochemistry, 116,* 1–9.

Sen, N., & Snyder, S. H. (2010). Protein modifications involved in neurotransmitter and gasotransmitter signaling. *Trends in Neuroscience, 33,* 493–502.

Stryer, L. (1995). *Biochemistry* (4th ed.). New York: Freeman.

Stryer, L., & Bourne, H. R. (1986). G proteins: A family of signal transducers. *Annual Review of Cell Biology, 2,* 391–419.

Svenningsson, P., Nishi, A., Fisone, G., Girault, J. A., Nairn, A. C., & Greengard, P. (2004). DARPP-32: An integrator of neurotransmission. *Annual Review of Pharmacology and Toxicology, 44,* 269–296.

Tanaka, C., & Nishizuka, Y. (1994). The protein kinase C family for neuronal signaling. *Annual Review of Neuroscience, 17,* 551–567.

Taussig, R., & Gilman, A. G. (1995). Mammalian membrane-bound adenylyl cyclases. *Journal of Biological Chemistry, 270,* 1–4.

Tjian, R., & Maniatis, T. (1994). Transcription activation: A complex puzzle with few easy pieces. *Cell, 77,* 5–8.

Wayman, G. A., Lee, Y. S., Tokumitsu, H., Silva, A. J., Silva, A., & Soderling, T. R. (2008). Calmodulin-kinases: Modulators of neuronal development and plasticity. *Neuron, 59,* 914–931.

Suggested Readings

Coba, M. P., Pocklington, A. J., Collins, M. O., Kopanitsa, M. V., Uren, R. T., Swamy, S., Croning, M. D. R., Choudhary, J. S., & Grant, S. G. N. (2009). Neurotransmitters drive combinatorial multistate postsynaptic density networks. *Science signaling, 2,* ra19.

Greengard, P. (2001). The neurobiology of slow synaptic transmission. *Science, 294,* 1024–1030.

Hagenston, A. M., & Bading, H. (2011). Calcium signaling in synapse-to-nucleus communication. *Cold Spring Harbor Perspectives in Biology*, Doi:10.1101/cshperspect.a004564.

Rellos, P., Pike, A. C. W., Niesen, F. H., Salah, E., Lee, W. H., von Delft, F., & Knapp, S. (2010). Structure of the CaMKIIδ/calmodulin complex reveals the molecular mechanism of CaMKII kinase activation. *PLoS Biology, 8,* e1000426.

Virshup, D. M., & Shenolikar, S. (2009). From promiscuity to precision: protein phosphatases get a makeover. *Molecular Cell, 33,* 537–545.

Howard Schulman

This page intentionally left blank

CHAPTER

10

Postsynaptic Potentials and Synaptic Integration

The study of synaptic transmission in the central nervous system (CNS) provides an opportunity to learn more about the diversity and richness of mechanisms underlying this process and to learn how some of the fundamental signaling properties of the nervous system, such as action potentials and synaptic potentials, work together to process information and generate behavior.

Postsynaptic potentials (PSPs) in the CNS can be divided into two broad classes on the basis of mechanisms and, generally, duration of these potentials. One class is based on the direct binding of a transmitter molecule(s) with a receptor-channel complex; these receptors are ionotropic. The structure of these receptors is discussed in detail in Chapter 8. The resulting PSPs are generally short-lasting and thus sometimes are called fast PSPs; they have also been referred to as "classical" because they were the first synaptic potentials to be recorded in the CNS (Eccles, 1964; Spencer, 1977). The duration of a typical fast PSP is about 20 ms.

The other class of PSPs is based on the indirect effect of a transmitter molecule(s) binding with a receptor. The receptors that produce these PSPs are metabotropic. As discussed in Chapter 8, the receptors activate G proteins (G-protein-coupled receptors; GPCRs) that affect the channel either directly or through additional steps in which the level of a second messenger is altered. The changes in membrane potential produced by metabotropic receptors can be long-lasting and therefore are called slow PSPs. The mechanisms for fast PSPs mediated by ionotropic receptors are considered first.

IONOTROPIC RECEPTORS: MEDIATORS OF FAST EXCITATORY AND INHIBITORY SYNAPTIC POTENTIALS

The Stretch Reflex Is Useful to Examine the Properties and Functional Consequences of Ionotropic PSPs

The stretch reflex, one of the simpler behaviors mediated by the central nervous system, is a useful example with which to examine the properties and functional consequences of ionotropic PSPs. The tap of a neurologist's hammer to a ligament elicits a reflex extension of the leg, as illustrated in Figure 10.1. The brief stretch of the ligament is transmitted to the extensor muscle and is detected by specific receptors in the muscle and ligament (Chapter 28). Action potentials initiated in the stretch receptors are propagated to the spinal cord by afferent fibers (Chapter 28). The receptors are specialized regions of sensory neurons with somata located in the dorsal root ganglia just outside the spinal column. Axons of the afferents enter the spinal cord and make excitatory synaptic connections with at least two types of postsynaptic neurons. First, a synaptic connection is made to the extensor motor neuron. As the result of its synaptic activation, the motor neuron fires action potentials that propagate out of the spinal cord and ultimately invade the terminal regions of the motor axon at neuromuscular junctions. There, acetylcholine (ACh) is released, nicotinic ACh receptors are activated, an end plate potential

© 2013 Elsevier Inc. All rights reserved.

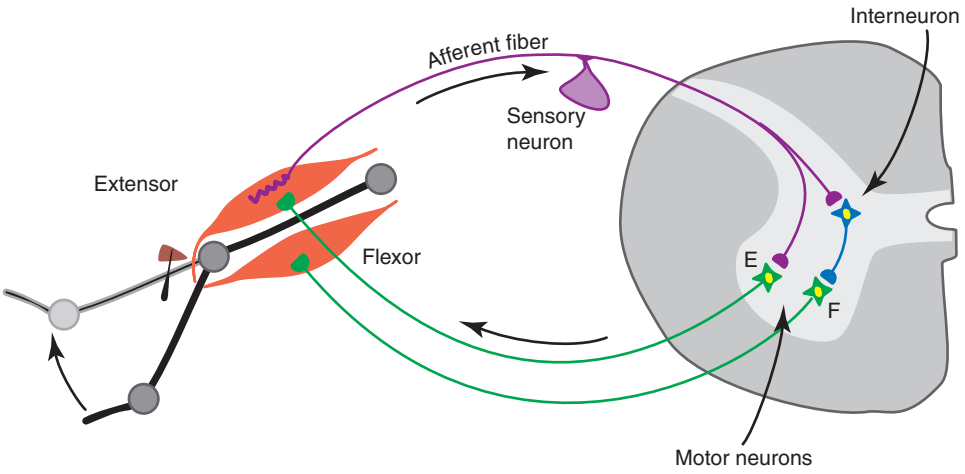


FIGURE 10.1 Features of the vertebrate stretch reflex. Stretch of an extensor muscle leads to the initiation of action potentials in the afferent terminals of specialized stretch receptors. The action potentials propagate to the spinal cord through afferent fibers (sensory neurons). The afferents make excitatory connections with extensor motor neurons (E). Action potentials initiated in the extensor motor neurons propagate to the periphery and lead to the activation and subsequent contraction of the extensor muscle. The afferent fibers also activate interneurons that inhibit the flexor motor neurons (F).

(EPP) is produced, an action potential is initiated in the muscle cell, and the muscle cell is contracted, producing the reflex extension of the leg. Second, a synaptic connection is made to another group of neurons called interneurons (nerve cells interposed between one type of neuron and another). The particular interneurons activated by the afferents are inhibitory interneurons because activation of these interneurons leads to the release of a chemical transmitter substance that inhibits the flexor motor neuron. This inhibition tends to prevent an uncoordinated (counteracting) movement (i.e., flexion) from occurring. The reflex system illustrated in Figure 10.1 also is known as the monosynaptic stretch reflex because this reflex is mediated by a single ("mono") excitatory synapse in the central nervous system. Spinal reflexes are described in greater detail in Chapter 28. Box 10.1 on optogenetics describes new techniques that can identify neuronal circuits.

Figure 10.2 illustrates procedures that can be used to experimentally examine some of the components of synaptic transmission in the reflex pathway for the stretch reflex. Intracellular recordings are made from one of the sensory neurons, the extensor and flexor motor neurons, and an inhibitory interneuron. Normally, the sensory neuron is activated by stretch to the muscle, but this step can be bypassed by simply injecting a pulse of depolarizing current of sufficient magnitude into the sensory neuron to elicit an action potential. The action potential in the sensory neuron leads to a potential change in the motor neuron known as an excitatory postsynaptic potential (EPSP; Fig. 10.2).

Mechanisms responsible for fast EPSPs mediated by ionotropic receptors in the CNS are fairly well known. Moreover, the ionic mechanisms for EPSPs in the CNS are essentially identical with the ionic mechanisms at the skeletal neuromuscular junction. Specifically, the transmitter substance released from the presynaptic terminal (Chapters 7 and 8) diffuses across the synaptic cleft, binds to specific receptor sites on the postsynaptic membrane (Chapter 8), and leads to a simultaneous increase in permeability to Na^+ and K^+, which makes the membrane potential move toward a value of about 0 mV. However, the processes of synaptic transmission at the sensory neuron–motor neuron synapse and the motor neuron–skeletal muscle synapse differ in two fundamental ways: in the transmitter used and in the amplitude of the EPSP. The transmitter substance at the neuromuscular junction is ACh, whereas that released by the sensory neurons is an amino acid, probably glutamate. Indeed, glutamate is the most common transmitter that mediates excitatory actions in the CNS. The amplitude of the postsynaptic potential at the neuromuscular junction is about 50 mV; consequently, each PSP depolarizes the postsynaptic cell beyond threshold so there is a one-to-one relation between an action potential in the spinal motor neuron and an action potential in the skeletal muscle cell. Indeed, the EPSP must depolarize the muscle cell by only about 30 mV to initiate an action potential, allowing a safety factor of about 20 mV. In contrast, the EPSP in a spinal motor neuron produced by an action potential in an afferent fiber has an amplitude of only about 1 mV. The mechanisms by which these small PSPs can trigger an action potential in the postsynaptic neuron are discussed in a later section of this chapter and in Chapter 11.

BOX 10.1

OPTOGENETICS

The brain's complexity is evident in its hundreds (at least) of neuronal cell types, each with different cellular morphologies, gene expression profiles, patterns of synaptic connectivity, and electrophysiological properties. Understanding how neural activity underlies the sophisticated computations and behavior produced by a brain—and how it is impaired in disease states—requires the ability to dissect the brain's complexity with the appropriate level of precision. The advent of optogenetics (light-activated genetically encoded proteins to manipulate cellular function) enables the nervous system to be dissected at an unparalleled level of specificity. This technology has ushered in a new era of neuroscience research, shedding light on aspects of brain function that less than a decade ago were wholly inaccessible.

One of the most powerful optogenetic technologies is the expanding set of single-component optogenetic tools developed from naturally occurring microbial rhodopsins. Like the rhodopsin underlying vision in the retina, the microbial rhodopsins are trans-membrane proteins which bind to the light-sensitive molecule retinal. Retinal undergoes a significant conformational change in the presence of light, which in turn catalyzes conformational changes in rhodopsin. Unlike rhodopsin in the eye, which is a G-protein-coupled receptor, many microbial rhodopsins are channels or pumps. The significance of this fact for neuroscience is that a single opsin protein (along with its bound organic co-factor retinal) fluxes ions in response to light, enabling the required functionality to be encoded in a single gene (i.e., a single-component tool).

The diverse optogenetic toolbox now enables powerful experimental control over neural activity. Channelrhodopsin-2 (ChR2) which is naturally found in the green algae *Chlamydomonas reinhardtii* was the first single-component optogenetic tool to be expressed in neurons, and exemplifies optogenetic capabilities. ChR2 is a non-selective cation channel that opens within microseconds following brief exposure to blue light (~470 nm). When expressed in neurons and activated by light, the flux of positive ions into the cell rapidly and reliably produces an action potential. Thus, using ChR2, an experimenter can elicit a particular desired pattern of neural firing simply by delivering the corresponding pattern of light. Such high-fidelity and temporally precise control over the neural code is unprecedented. A diverse set of optogenetic tools with complementary capabilities have arisen from multiple lines of research, including from genomic searches for novel naturally occuring microbial opsins, experimentation with chimeric proteins constructed from multiple rhodopsins, and rationally designed mutations of specific amino acids in wildtype (WT) rhodopsins. Today a wide variety of optogenetic tools exist which are faster, more reliable, more potent, and more light sensitive than WT ChR2 (e.g., ChR2-H134R, ChETA, C1V1); some of these new tools can even produce stable, long-lasting depolarizing currents for up to tens of minutes after a single brief light flash, and can be precisely activated or inactivated with different wavelengths of light (e.g., SFO, SSFO). Rhodopsins have also been produced that are sensitive to different wavelengths of light (e.g., blue, green, red), enabling simultaneous independent optogenetic control of different neural populations (e.g., VChR1, C1V1). Microbial opsins which hyperpolarize cells and inhibit action potentials such as the chloride pump halorhodopsin (NpHR) found in the halobacterium *Natronomonas pharaonis* have been discovered and methodically enhanced, thus enabling bi-directional control of neural activity (e.g., eNpHR3.0, eArch). Finally, there are tools which enable the optical control of intracellular signaling through G-protein-mediated molecular cascades (optoXRs). Given how rapidly these tools have been described and elaborated upon by a number of productive laboratories, the growing capabilities of optogenetic proteins will likely continue to expand.

It is important to remember that the power of optogenetics is limited by the degree to which these proteins can be selectively expressed and activated in specific neuronal cell types *in vitro* and *in vivo*. However, since the opsins are genetically encoded proteins, scientists have an arsenal of gene therapy and molecular biology techniques available to achieve selective expression in neural tissues. These include *in vitro* cell culture transfection, *in utero* electroporation, knock-in, knock-out and transgenic animals, promoter/enhancer manipulations, viral vectors, and inducible expression tools such as those involving Cre recombinase, many of which can be used in combination. Experiments *in vitro* can involve light delivery from mercury arc lamps, lasers, or light-emitting diodes (LEDs), and 2-photon optogenetic activation of neurons has also been achieved. In addition, both anesthetized and awake animals can be optogenetically manipulated; for *in vivo* stimulation laser light is typically delivered through <1 mm optical fibers, which are stereotaxically targeted to specific brain regions. Using these various gene therapy and optical approaches, neuroscientists have used optogenetic technologies in a variety of tissues and awake behaving organisms, including *Caenorhabditis elegans*, *Drosophila*, zebrafish, mouse, rat, macaque monkey *in vivo*, and *in vitro* human tissue. The era of optogenetic neuroscience research has already made a significant

BOX 10.1 (*cont'd*)

impact, but as the development and application of optogenetics continues additional merits of this technology are likely to be realized.

Ryan Fox Squire and Karl Deisseroth

Further Reading

Deisseroth, K. (2010). Optogenetics: Controlling the brain with light [Extended Version]: Scientific American. *Scientific American* (web exclusive). http://www.scientificamerican.com/article.cfm?id=optogenetics-controlling.

Fenno, L., Yizhar, O., & Deisseroth, K. (2011). The development and application of optogenetics. *Annual Review of Neuroscience*, *34*, 389–412.

Luo, L., Callaway, E. M., & Svoboda, K. (2008). Genetic dissection of neural circuits. *Neuron*, *57*(5), 634–660.

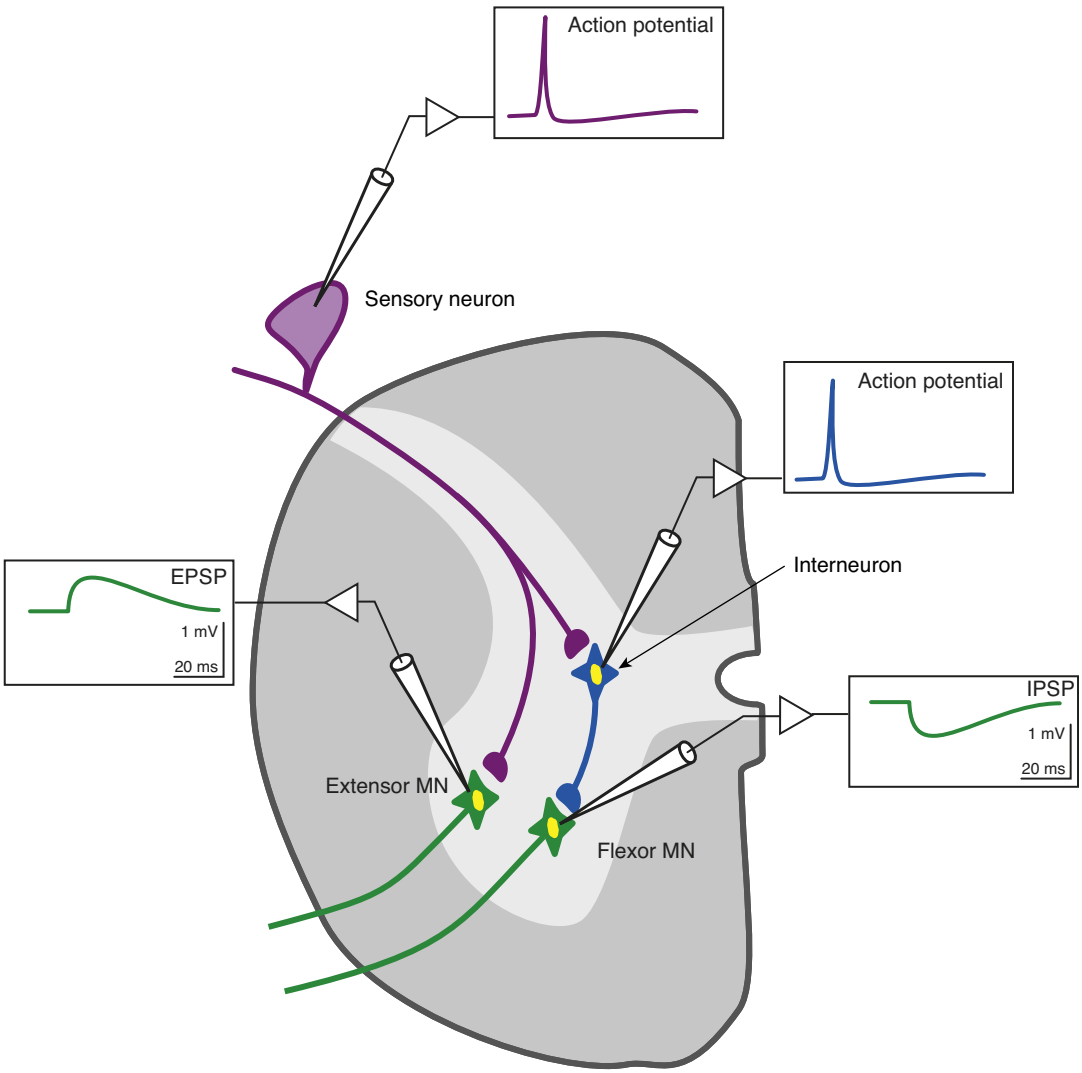


FIGURE 10.2 Excitatory (EPSP) and inhibitory (IPSP) postsynaptic potentials in spinal motor neurons. Idealized intracellular recordings from a sensory neuron, interneuron, and extensor and flexor motor neurons (MNs). An action potential in the sensory neuron produces a depolarizing response (an EPSP) in the extensor motor neuron. An action potential in the interneuron produces a hyperpolarizing response (an IPSP) in the flexor motor neuron.

Macroscopic Properties of PSPs Are Determined by the Nature of Gating and Ion-Permeation Properties of Single Channels

Patch-Clamp Techniques

Patch-clamp techniques (Hamill, Marty, Neher, Sakmann, & Sigworth, 1981), with which current flowing through single isolated receptors can be measured directly, can be sources of insight into both the ionic mechanisms and the molecular properties of PSPs mediated by ionotropic receptors. This approach was pioneered by Erwin Neher and Bert Sakman in the 1970s and led to their being awarded the Nobel Prize in Physiology or Medicine in 1991.

Figure 10.3A illustrates an idealized experimental arrangement of an "outside-out" patch recording of a single ionotropic receptor. The patch pipette contains a solution with an ionic composition similar to that of the cytoplasm, whereas the solution exposed to the outer surface of the membrane has a composition similar to that of normal extracellular fluid. The electrical potential across the patch, and hence the transmembrane potential (V_m), is controlled by the patch-clamp amplifier. The extracellular (outside) fluid is considered "ground." Transmitter can be delivered by applying pressure to a miniature pipette filled with an agonist (in this case, ACh), and the current (I_m) flowing across the patch of membrane is measured by the patch-clamp amplifier (Fig. 10.3). Pressure in the pipette that contains ACh can be continuous, allowing a constant stream of ACh to contact the membrane, or can be applied as a short pulse to allow a precisely timed and discrete amount of ACh to contact the membrane. The types of recordings obtained from such an experiment are illustrated in the traces in Figure 10.3. In the absence of ACh, no current flows through the channel (Fig. 10.3A). When ACh is applied continuously, current flows across the membrane (through the channel), but the current does not flow continuously; instead, small step-like changes in current are observed (Fig. 10.3B). These changes represent the probabilistic (random) opening and closing of the channel.

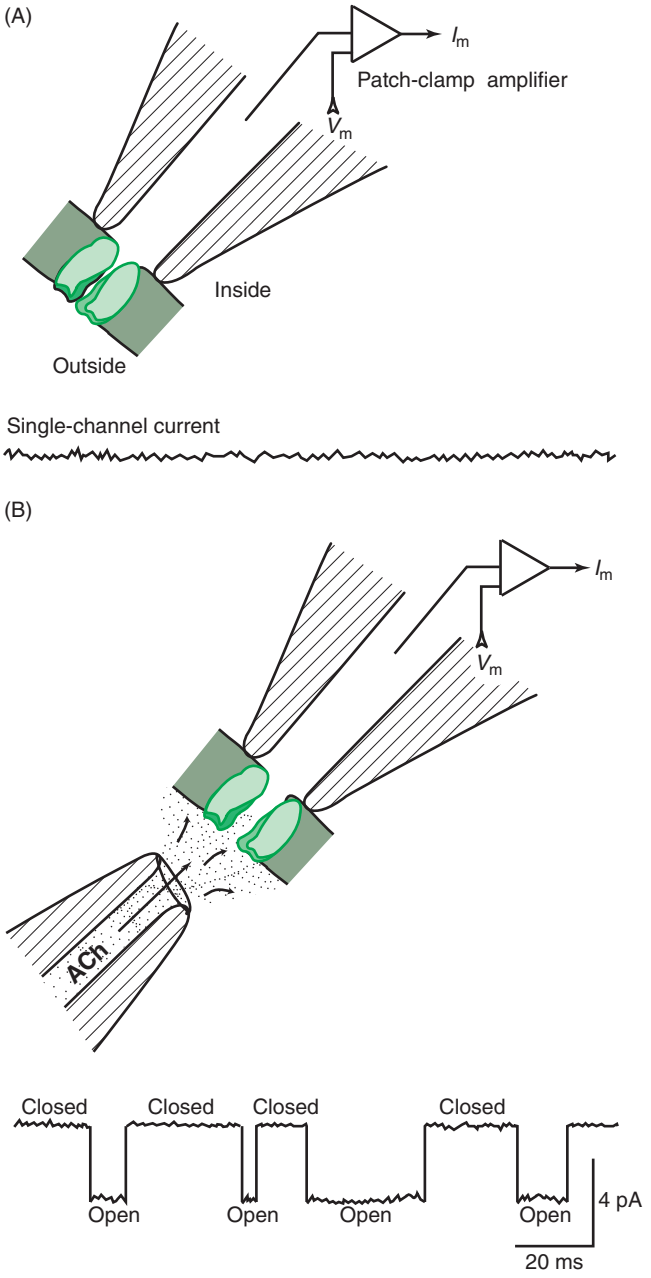


FIGURE 10.3 Single-channel recording of ionotropic receptors and their properties. (A) Experimental arrangement for studying properties of ionotropic receptors. (B) Idealized single-channel currents in response to application of ACh.

Channel Openings and Closings

As a result of the type of patch-recording techniques heretofore described, three general conclusions about the properties of ligand-gated channels can be drawn. First, ACh, as well as other transmitters that activate ionotropic receptors, causes the opening of individual ionic channels (for a channel to open, usually two molecules of transmitter must bind to the receptor). Second, when a ligand-gated channel opens, it does so in an all-or-none fashion. Increasing the concentration of transmitter in the ejection microelectrode does not increase the permeability (conductance) of the channel; it increases its probability (P) of being open. Third, the ionic current flowing through a single channel in its open state is extremely small (e.g., 10^{-12} A); as a result, current flowing through any single channel makes only a small contribution to the normal postsynaptic potential. Physiologically, when a larger region of the postsynaptic membrane, and thus more than one channel, is exposed to a released transmitter, the net conductance

of the membrane increases due to the increased probability that a larger population of channels will be open at the same time. The normal PSP, measured with standard intracellular recording techniques (e.g., Fig. 10.2), is then proportional to the sum of the currents that flow through these many individual open channels. The properties of voltage-sensitive channels (see Chapter 5) are similar in that they, too, open in all-or-none fashion, and, as a result, the net effect on the cell is due to the summation of currents flowing through many individual open ion channels. The two types of channels differ, however, in that one is opened by a chemical agent, whereas the other is opened by changes in membrane potential.

Statistical Analysis of Channel Gating and Kinetics of the PSP

The experiment illustrated in Figure 10.3B was performed with continuous exposure to ACh. Under such conditions, the channels open and close repeatedly. When ACh is applied by a brief pressure pulse to more accurately mimic the transient release from the presynaptic terminal, the transmitter commonly causes only a single opening of the channel before it diffuses away. A set of data similar to that shown in Figure 10.4A would be obtained if an ensemble of these openings were collected and aligned with the start of each opening. Each individual trace represents the response to each successive "puff" of ACh. Note that, among the responses, the duration of the opening of the channel varies considerably—from very short (less than 1 ms) to more than 5 ms. Moreover, channel openings are independent events. The duration of any one channel opening does not have any relation to the duration of a previous opening. Figure 10.4B illustrates a plot that is obtained by adding 1000 of these individual responses. Such an addition roughly simulates the conditions under which a transmitter released from a presynaptic terminal leads to the near simultaneous activation of many single channels in the postsynaptic membrane. (Note that the addition of 1000 channels would produce a synaptic current equal to about 4 nA.) This simulation is valid given the assumption that the statistical properties of a single channel over time are the same as the statistical properties of the ensemble at one instant of time (i.e., an ergodic process). The ensemble average can be fit with an exponential function with a decay time constant of 2.7 ms. An additional observation (discussed later) is that the value of the time constant is equal to the mean duration of the channel openings. The curve in Figure 10.4B is an indication of the probability that a channel will remain open for various times, with a high probability for short times and a low probability for long times.

The ensemble average of single-channel currents (Fig. 10.4B) roughly accounts for the time course of the

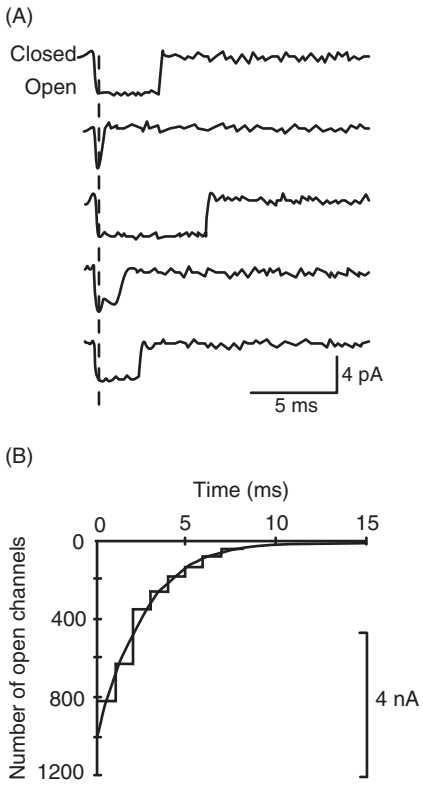


FIGURE 10.4 Determination of the shape of the postsynaptic response from single-channel currents. (A) Each trace represents the response of a single channel to a repetitively applied puff of transmitter. Traces are aligned with the beginning of the channel opening (dashed line). (B) The addition of 1000 of the individual responses. If a current equal to 4 pA were generated by the opening of a single channel, then a 4-nA current would be generated by 1000 channels opening at the same time. Data are fitted with an exponential function having a time constant equal to 1/a (see text). *Reprinted with permission from Sakmann (1992). American Association for the Advancement of Science, © 1992 The Nobel Foundation.*

EPSP. However, note that the time course of the aggregate synaptic current can be somewhat faster than that of the excitatory postsynaptic potential in Figure 10.2. This difference is due to charging of the membrane capacitance by a rapidly changing synaptic current. Because the single-channel currents were recorded with the membrane voltage clamped, the capacitive current $[I_c = C_m * (dV/dt)]$ is zero. In contrast, for the recording of the postsynaptic potential in Figure 10.2, the membrane was not voltage clamped, and therefore as the voltage changes (i.e., dV/dt), some of the synaptic current charges the membrane capacitance (see Eq. 10.7).

Analytical expressions that describe the shape of the ensemble average of the open lifetimes and the mean open lifetime can be derived by considering that single-channel opening and closing is a stochastic process (Johnston & Wu, 1995; Sakmann, 1992). Relations are formalized to describe the likelihood (probability)

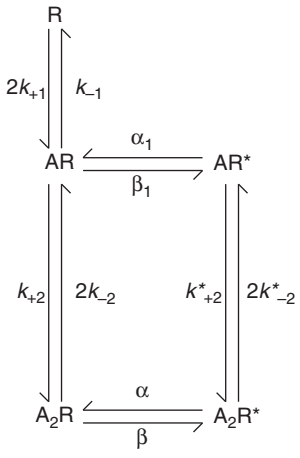
of a channel being in a certain state. Consider the following two-state reaction scheme:

$$C \underset{\alpha}{\overset{\beta}{\rightleftharpoons}} O$$

In this scheme, α represents the rate constant for channel closing and β the rate constant for channel opening. The scheme can be simplified further if we consider a case in which the channel has been opened by the agonist and the agonist is removed instantaneously. A channel so opened (at time 0) will then close after a certain random time (Fig. 10.4). It can be shown that the mean open time = $1/\alpha$ (Johnston & Wu, 1995; Sakmann, 1992).

Gating Properties of Ligand-Gated Channels

Although statistical analysis can be a valuable source of insight into the statistical nature of the gating process and the molecular determinants of the macroscopic postsynaptic potential, the description in the preceding section is a simplification of the actual processes. Specifically, a more complete description must include the kinetics of receptor binding and unbinding and the determinants of the channel opening, as well as the fact that channels display rapid transitions between open and closed states during a single agonist receptor occupancy. Thus, the open states illustrated in Figures 10.3B and 10.4A represent the period of a burst of extremely rapid openings and closings. If the bursts of rapid channel openings and closings are thought of, and behave functionally, as a single continuous channel opening, the formalism developed in the preceding section is a reasonable approximation for many ligand-gated channels. Nevertheless, a more complex reaction scheme is necessary to quantitatively explain available data. Such a scheme would include the following states,



where R represents the receptor, A the agonist, and the α, β, and k values the forward and reverse rate constants for the various reactions. A_2R^* represents a channel opened as a result of the binding of two agonist molecules. The asterisk indicates an open channel. Note that the lower part of the reaction scheme is equivalent to the one developed earlier; that is,

$$C \underset{\alpha}{\overset{\beta}{\rightleftharpoons}} O$$

With the use of probability theory, equations describing transitions between the states can be determined. The approach is identical to that used in the simplified two-state scheme. However, the mathematics and analytical expressions are more complex. For some receptors, additional states must be represented. For example, as described in Chapter 8, some ligand-gated channels exhibit a process of desensitization in which continued exposure to a ligand results in channel closure.

Null (Reversal) Potential and Slope of I-V Relations

What ions are responsible for the synaptic current that produces the EPSP? Early studies of the ionic mechanisms underlying the EPSP at the skeletal neuromuscular junction yielded important information. Specifically, voltage-clamp and ion-substitution experiments indicated that the binding of transmitter to receptors on the postsynaptic membrane led to a simultaneous increase in Na^+ and K^+ permeability that depolarized the cell toward a value of about 0 mV (Fatt & Katz, 1951; Takeuchi & Takeuchi, 1960). These findings are applicable to the EPSP in a spinal motor neuron produced by an action potential in an afferent fiber and have been confirmed and extended at the single-channel level.

Figure 10.5 illustrates the type of experiment in which the analysis of single-channel currents can be a source of insight into the ionic mechanisms of EPSPs. A transmitter is delivered to the patch while the membrane potential is varied systematically (Fig. 10.5A). In the upper trace, the patch potential is −40 mV. The ejection of transmitter produces a sequence of channel openings and closings, the amplitudes of which are constant for each opening (i.e., about 4 pA). Now consider the case in which the transmitter is applied when the potential across the patch is −20 mV. The frequency of the responses, as well as the mean open lifetimes, is about the same as when the potential was at −40 mV, but now the amplitude of the single-channel currents is decreased uniformly. Even more interesting, when the patch is depolarized artificially to a value of about 0 mV, an identical puff of transmitter produces no current in the patch. If the patch potential is depolarized to a value of about 20 mV and the puff is delivered again, openings are again observed, but the flow of current through the channel is reversed in sign; a series of

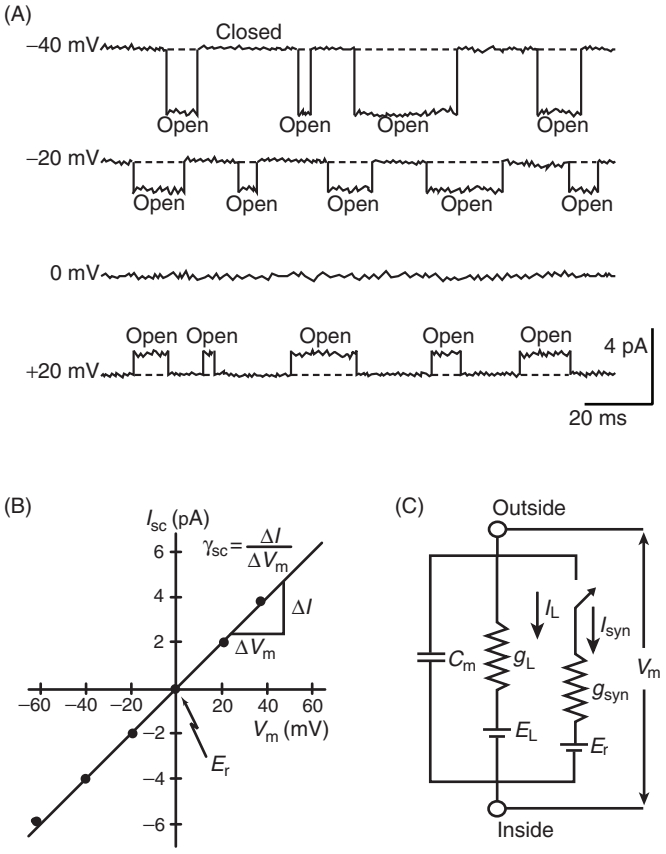


FIGURE 10.5 Voltage dependence of the current flowing through single channels. (A) Idealized recording of an ionotropic receptor in the continuous presence of agonist. (B) I–V relation of the channel in A. (C) Equivalent electrical circuit of a membrane containing that channel. γ_{SC}, single-channel conductance; I_L, leakage current; I_{SC}, single-channel current; g_L, leakage conductance; g_{syn}, macroscopic synaptic conductance; E_L, leakage battery; E_r, reversal potential.

upward deflections indicate outward single-channel currents. In summary, there are downward deflections (inward currents) when the membrane potential is at −40 mV, no deflections (currents) when the membrane is at 0 mV, and upward deflections (outward currents) when the membrane potential is moved to 20 mV.

The simple explanation for these results is that no matter what the membrane potential, the effect of the transmitter binding with receptors is to produce a permeability change that tends to move the membrane potential toward 0 mV. If the membrane potential is more negative than 0 mV, an inward current is recorded. If the membrane potential is more positive than 0mV, an outward current is recorded. If the membrane potential is at 0 mV, there is no deflection because the membrane potential is already at 0 mV. At 0 mV, the channels are opening and closing as they always do in response to the agonist, but there is no net movement of ions through them. This 0-mV level is known as the synaptic null potential or reversal potential because it is the potential at which the sign of the synaptic current reverses. The fact that the experimentally determined reversal potential equals the calculated value obtained by using the Goldman-Hodgkin-Katz (GHK) equation (Chapter 5) provides strong support for the theory that the EPSP is due to the opening of channels that have equal permeabilities to Na^+ and K^+. Ion-substitution experiments also confirm this theory. Thus, when the concentration of Na^+ or K^+ in the extracellular fluid is altered, the value of the reversal potential shifts in a way predicted by the GHK equation. (Some other cations, such as Ca^{2+}, also permeate these channels, but their permeability is low compared with that of Na^+ and K^+.)

Different families of ionotropic receptors have different reversal potentials because each has unique ion selectivity. In addition, it should now be clear that the sign of the synaptic action (excitatory or inhibitory) depends on the value of the reversal potential relative to the resting potential. If the reversal potential of an ionotropic receptor channel is more positive than the resting potential, opening of that channel will lead to depolarization (i.e., an EPSP). In contrast, if the reversal potential of an ionotropic receptor channel is more negative than the resting potential, opening of that channel will lead to hyperpolarization; that is, an inhibitory postsynaptic potential (IPSP), which is the topic of a later section in this chapter.

Plotting the average peak value of single-channel currents (I_{sc}) versus the membrane potential (transpatch potential) at which they are recorded (Fig. 10.5B) can be a source of quantitative insight into the properties of the ionotropic receptor channel. Note that the current-voltage (*I-V*) relation is linear; it has a slope, the value of which is the single-channel conductance, and an intercept at 0 mV. This linear relation can be put in the form of Ohm's law ($I = G * \Delta V$). Thus,

$$I_{sc} = \gamma_{sc} * (V_m - E_r), \tag{10.1}$$

where γ_{sc} is the single-channel conductance and E_r is the reversal potential (here, 0 mV).

Summation of Single-Channel Currents

We now know that the sign of a synaptic action can be predicted by knowledge of the relation between the resting potential (V_m) and the reversal potential (E_r), but how can the precise amplitude be determined? The answer to this question lies in understanding the relation between synaptic conductance and extra synaptic conductances. These interactions can be rather complex (see Chapter 11), but some initial understanding can be obtained by analyzing an electrical equivalent circuit for these two major conductance branches. We first need to move from a consideration of single-channel conductances and currents to that of

macroscopic conductances and currents. The postsynaptic membrane contains thousands of any one type of ionotropic receptor, and each of these receptors could be activated by transmitter released by a single action potential in a presynaptic neuron. Because conductances in parallel add, the total conductance change produced by their simultaneous activation would be

$$g_{syn} = \gamma_{sc} P * N, \tag{10.2}$$

where γ_{sc}, as before, is the single-channel conductance, P is the probability of opening of a single channel (controlled by the ligand), and N is the total number of ligand-gated channels in the postsynaptic membrane. The macroscopic postsynaptic current produced by the transmitter released by a single presynaptic action potential can then be described by

$$I_{syn} = g_{syn} * (V_m - E_r). \tag{10.3}$$

Equation 10.3 can be represented physically by a voltage (V_m) measured across a circuit consisting of a resistor (g_{syn}) in series with a battery (E_r). An equivalent circuit of a membrane containing such a conductance is illustrated in Figure 10.5C. Also included in this circuit are a membrane capacitance (C_m), a resistor representing the leakage conductance (g_L), and a battery (E_L) representing the leakage potential. (Voltage-dependent Na^+, Ca^{2+}, and K^+ channels that contribute to the generation of the action potential have been omitted for simplification.)

The simple circuit allows the simulation and further analysis of the genesis of the PSP. Closure of the switch simulates the opening of the channels by transmitter released from some presynaptic neuron (i.e., a change in P of Eq. (10.2) from 0 to 1). When the switch is open (i.e., no agonist is present and the ligand-gated channels are closed), the membrane potential (V_m) is equal to the value of the leakage battery (E_L). Closure of the switch (i.e., the agonist opens the channels) tends to polarize the membrane potential toward the value of the battery (E_r) in series with the synaptic conductance. Although the effect of the channel openings is to depolarize the postsynaptic cell *toward* E_r (0 mV), this value is never achieved because ligand-gated receptors are only a small fraction of the ion channels in the membrane. Other channels (such as the leakage channels, which are not affected by the transmitters) tend to hold the membrane potential at E_L and prevent the membrane potential from reaching the 0-mV level. In terms of the equivalent electrical circuit (Fig. 10.5C), g_L is much greater than g_{syn}.

An analytical expression that can be a source of insight into the production of an EPSP by the engagement of a synaptic conductance can be derived by examining the current flowing in each of the two conductance branches of the circuit in Figure 10.5C. As shown previously (Eq. 10.3), current flowing in the branch representing the synaptic conductance is equal to

$$I_{syn} = g_{syn} * (V_m - E_r).$$

Similarly, the current flowing through the leakage conductance is equal to

$$I_L = g_L * (V_m - E_L). \tag{10.4}$$

By conservation of current, the two currents must be equal and opposite. Therefore,

$$g_{syn} * (V_m - E_r) = -g_L * (V_m - E_L).$$

Rearranging and solving for V_m, we obtain

$$V_m = \frac{g_{syn} E_r + g_L E_L}{g_{syn} + g_L}. \tag{10.5}$$

Note that when the synaptic channels are closed (i.e., switch open), g_{syn} is 0 and

$$V_m = E_L.$$

Now consider the case of ligand-gated channels being opened by the release of transmitter from a presynaptic neuron (i.e., switch closed) and a neuron with g_L = 10 nS, E_L = −60 mV, g_{syn} = 0.2 nS, and E_r = 0 mV. Then

$$V_m = \frac{(0.2\times10^{-9} * 0) + (10\times10^{-9} * -60)}{10.2\times10^{-9}}.$$
$$= -59\,\text{mV}$$

Thus, as a result of the closure of the switch, the membrane potential has changed from its initial value of −60 mV to a new value of −59 mV; that is, an EPSP of 1 mV has been generated.

The preceding analysis ignored membrane capacitance (C_m), the charging of which makes the synaptic potential slower than the synaptic current. Thus, a more complete analytical description of the postsynaptic factors underlying the generation of a PSP must account for the fact that some of the synaptic current will flow into the capacitive branch of the circuit. Again, by conservation of current, the sum of the currents in the three branches must equal 0. Therefore,

$$0 = C_m \frac{dV_m}{dt} + I_L + I_{syn}, \tag{10.6}$$

$$0 = C_m \frac{dV_m}{dt} + g_L * (V_m - E_L) g_{syn}(t) * (V_m - E_r), \tag{10.7}$$

where C_m (dV_m/dt) is the capacitative current.

By solving for V_m and integrating the differential equation, we can determine the magnitude and time

course of a PSP. An accurate description of the kinetics of the PSP requires that the simple switch closure (all-or-none engagement of the synaptic conductance) be replaced with an expression [$g_{syn}(t)$] that describes the dynamics of the change in synaptic conductance with time.

Nonlinear I-V Relations of Some Ionotropic Receptors

For many PSPs mediated by ionotropic receptors, the current-voltage relation of the synaptic current is linear or approximately linear (Fig. 10.5B). Such ohmic relations are typical of nicotinic ACh channels and AMPA (alpha-amino-3-hydroxyl-5-methyl-4-isoxazolepropionate) glutamate channels (as well as many receptors mediating IPSPs). The linear I–V relation is indicative of a channel whose conductance is not affected by the potential across the membrane. Such linearity should be contrasted with the steep voltage dependency of the conductance of channels underlying the initiation and repolarization of action potentials (Chapter 5).

NMDA (*N*-methyl-D-aspartate) glutamate channels are a class of ionotropic receptors that have nonlinear current-voltage relations. At negative potentials, the channel conductance is low even when glutamate is bound to the receptor. As the membrane is depolarized, conductance increases and current flowing through the channel increases, resulting in the type of I–V relation illustrated in Figure 10.6A. This nonlinearity is represented by an arrow through the resistor representing this synaptic conductance in the equivalent circuit of Figure 10.6B. The nonlinear I–V relation of the NMDA receptor can be explained by a voltage-dependent block of the channel by Mg^{2+} (Fig. 10.7B). At normal values of the resting potential, the pore of the channel is blocked by Mg^{2+}. Thus, even when glutamate binds to the receptor (Fig. 10.7B), the blocked channel prevents ionic flow (and an EPSP). The block can be relieved by depolarization, which displaces Mg^{2+} from the pore (Fig. 10.7B). When the pore is unblocked, cations (i.e., Na^+, K^+, and Ca^{2+}) can flow readily through the channel, and this flux is manifested in the linear part of the I–V relation (Fig. 10.6A). AMPA channels (Fig. 10.7A) are not blocked by Mg^{2+} and have linear I–V relations (Fig. 10.5B).

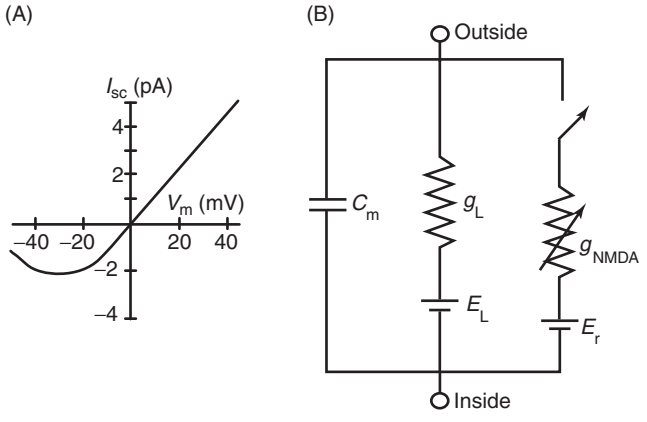


FIGURE 10.6 (A) I–V relation of the NMDA receptor. (B) Equivalent electrical circuit of a membrane containing NMDA receptors.

Inhibitory Postsynaptic Potentials Decrease the Probability of Cell Firing

Some synaptic events decrease the probability of generating action potentials in the postsynaptic cell. Potentials associated with these actions are called inhibitory postsynaptic potentials. Consider the inhibitory interneuron illustrated in Figure 10.2. Normally, this interneuron is activated by summating EPSPs from converging afferent fibers. These EPSPs summate in space and time such that the membrane potential of the interneuron reaches threshold and fires an action potential. This step can be bypassed by artificially depolarizing the interneuron to initiate an action potential. The consequences of that action potential from the point of view of the flexor motor neuron are illustrated in Figure 10.2. The action potential in the interneuron produces a transient increase in the membrane potential of the motor neuron. This transient hyperpolarization (the IPSP) looks very much like the EPSP, but it is reversed in sign.

What are the ionic mechanisms for these fast IPSPs and what is the transmitter substance? Because the membrane potential of the flexor motor neuron is about −65 mV, one might expect an increase in the conductance to some ion (or ions) with an equilibrium potential (reversal potential) more negative than −65 mV. One possibility is K^+. Indeed, the K^+ equilibrium potential in spinal motor neurons is about −80 mV; thus, a transmitter substance that produced a selective increase in K^+ conductance would lead to an IPSP. The K^+-conductance increase would move the membrane potential from −65 mV toward the K^+ equilibrium potential of −80 mV. Although an increase in K^+ conductance mediates IPSPs at some inhibitory synapses (see later), it does not at the synapse between the inhibitory interneuron and the spinal motor neuron. At this particular synapse, the IPSP seems to be due to a selective increase in Cl^- conductance. The equilibrium potential for Cl^- in spinal motor neurons is about −70 mV. Thus, the transmitter substance released by the inhibitory neuron diffuses across the cleft and interacts with receptor sites on the postsynaptic membrane. These receptors are normally closed, but when opened they become selectively permeable to Cl^-. As a result of the increase in Cl^- conductance, the membrane potential moves from a resting value of −65 mV toward the Cl^- equilibrium potential of −70 mV.

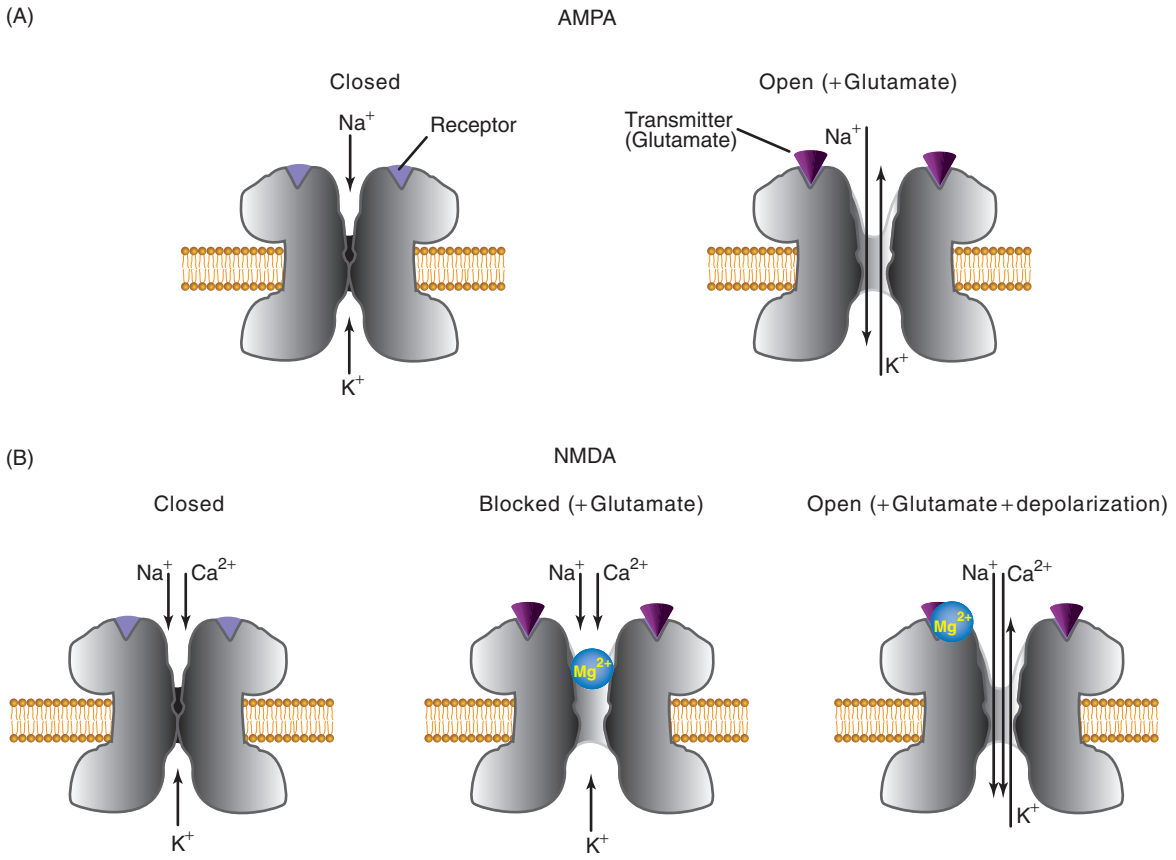


FIGURE 10.7 Features of AMPA and NMDA glutamate receptors. (A) AMPA receptors: (left) in the absence of agonist, the channel is closed, and (right) glutamate binding leads to channel opening and an increase in Na^+ and K^+ permeability. AMPA receptors that contain the GluR2 subunit are impermeable to Ca^{2+}. (B) NMDA receptors: (left) in the absence of agonist, the channel is closed; (middle) the presence of agonist leads to a conformational change and channel opening, but no ionic flux occurs because the pore of the channel is blocked by Mg^{2+}; and (right) in the presence of depolarization, the Mg^{2+} block is removed and the agonist-induced opening of the channel leads to changes in ion flux (including Ca^{2+} influx into the cell).

As in the sensory neuron–spinal motor neuron synapse, the transmitter substance released by the inhibitory interneuron in the spinal cord is an amino acid, but in this case the transmitter is glycine. The toxin strychnine is a potent antagonist of glycine receptors. Although glycine originally was thought to be localized to the spinal cord, it is also found in other regions of the nervous system. The most common transmitter associated with inhibitory actions in many areas of the brain is γ-aminobutyric acid (GABA; see Chapter 5).

GABA receptors are divided into three major classes: $GABA_A$, $GABA_B$, and $GABA_C$ (Billinton et al., 2001; Bormann & Fiegenspan, 1995; Bowery, 1993; Cherubini & Conti, 2001; Gage, 1992; Moss & Smart, 2001). As discussed in Chapter 7, $GABA_A$ receptors are ionotropic receptors, and, like glycine receptors, binding of transmitter leads to an increased conductance to Cl^-, which produces an IPSP. $GABA_A$ receptors are blocked by bicuculline and picrotoxin. A particularly striking aspect of $GABA_A$ receptors is their modulation by anxiolytic benzodiazepines. Figure 10.8 illustrates the response of a neuron to GABA before and after treatment with diazepam (Bormann, 1988). In the presence of diazepam, the response is potentiated greatly. In contrast to $GABA_A$ receptors that are pore-forming channels, $GABA_B$ receptors are G-protein coupled (see also Chapter 8). $GABA_B$ receptors can be coupled to a variety of different effector mechanisms in different neurons. These mechanisms include decreases in Ca^{2+} conductance, increases in K^+ conductance, and modulation of voltage-dependent A-type K^+ current. In hippocampal pyramidal neurons, the $GABA_B$-mediated IPSP is due to an increase in K^+ conductance. Baclofen is a potent agonist of $GABA_B$ receptors, whereas phaclofen is a selective antagonist. $GABA_C$ receptors are pharmacologically distinct from $GABA_A$ and $GABA_B$ receptors and are found predominantly in the vertebrate retina. $GABA_C$ receptors, like $GABA_A$ receptors, are Cl^- selective pores.

Ionotropic receptors that lead to the generation of IPSPs and ionotropic receptors that lead to the generation of EPSPs have biophysical features in common. Indeed, analyses of the preceding section are generally

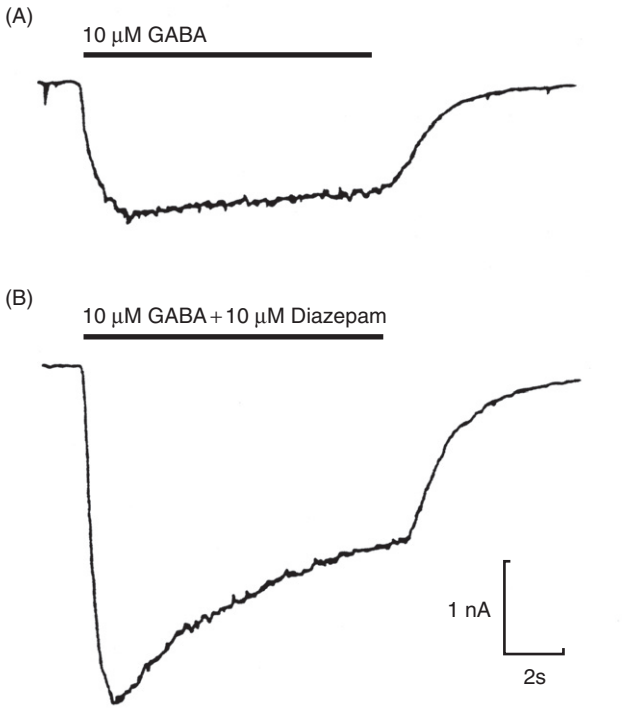


FIGURE 10.8 Potentiation of GABA responses by benzodiazepine ligands. (A) Brief application (bar) of GABA leads to an inward Cl^- current in a voltage-clamped spinal neuron. (B) In the presence of diazepam, the response is enhanced significantly. *From Bormann (1988).*

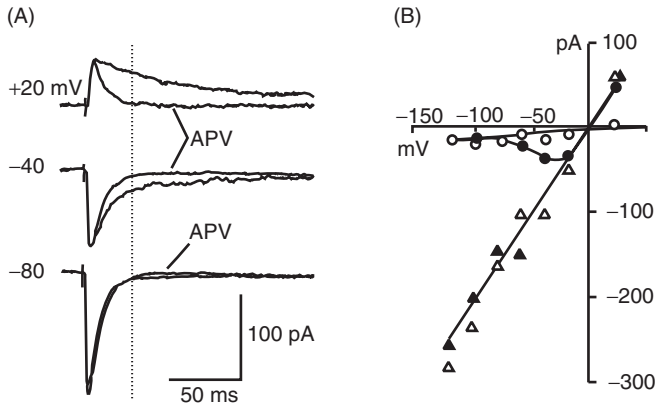


FIGURE 10.9 Dual-component glutamatergic EPSP. (A) The excitatory postsynaptic current was recorded before and during the application of APV at the indicated membrane potentials. (B) Peak current-voltage relations are shown before (▲) and during (△) the application of APV. Current-voltage relations measured 25 ms after the peak of the EPSC (dotted line in (A)); before (●) and during (○) application of APV are also shown. *Reprinted with permission from Hestrin et al. (1990).*

applicable. A quantitative understanding of the effects of the opening of glycine or $GABA_A$ receptors can be obtained by using the electrical equivalent circuit of Figure 10.5C and Eq. 10.5, with the values of g_{syn} and E_r appropriate for the respective ionotropic receptor. Interactions between excitatory and inhibitory conductances can be modeled by adding additional branches to the equivalent circuit (see Fig. 10.15D and Chapter 12).

Some PSPs Have More Than One Component

The transmitter released from a presynaptic terminal diffuses across the synaptic cleft, where it binds to ionotropic receptors. In many cases, the postsynaptic receptors are homogeneous. In other cases, the same transmitter activates more than one type of receptor. A major example of this type of heterogeneous postsynaptic action is the simultaneous activation by glutamate of NMDA and AMPA receptors on the same postsynaptic cell. Figure 10.9A illustrates such a dual-component glutamatergic EPSP in the CA1 region of the hippocampus. The cell is voltage clamped at various fixed holding potentials, and the macroscopic synaptic currents produced by activation of the presynaptic neurons are recorded. The experiment is performed in the presence and absence of the agent 2-amino-5-phosphonovalerate (APV), which is a specific blocker of NMDA receptors. When the cell is held at a potential of 20 or −40 mV, APV leads to a dramatic reduction of the late, but not the early, phase of the excitatory postsynaptic current (EPSC). In contrast, when the potential is held at −80 mV, the EPSC is unaffected by APV. These results indicate that the PSP consists of two components: an early AMPA-mediated component and a late NMDA-mediated component. In addition, results indicate that conductance of the non-NMDA component is linear, whereas conductance of the NMDA component is nonlinear. The I–V relations of the early (peak) and late (at approximately 25 ms) components of the EPSC are plotted in Figure 10.9B (Hestrin, Nicoll, Perkel, & Sah, 1990).

Dual-component IPSPs also are observed in the CNS, but here the transmitter (GABA) that mediates the inhibitory actions may be released from different neurons that converge on a common postsynaptic neuron. Stimulation of afferent pathways to the hippocampus results in an IPSP in a pyramidal neuron, which has a fast initial inhibitory phase followed by a slower inhibitory phase (Fig. 10.10). Application of $GABA_A$ antagonists blocks the early inhibitory phase, whereas the $GABA_B$ receptor antagonist phaclofen blocks the late inhibitory phase (not shown). Early and late IPSPs can also be distinguished based on their ionic mechanisms. Hyperpolarizing the membrane potential to −78 mV nulls the early response, but at this value of membrane potential, the late response is still hyperpolarizing (Figs. 10.10A and 10.10B). Hyperpolarizing the membrane potential to values more negative than −78 mV reverses the sign of the early response, but the slow response does not reverse until the membrane is made more negative than about −100 mV

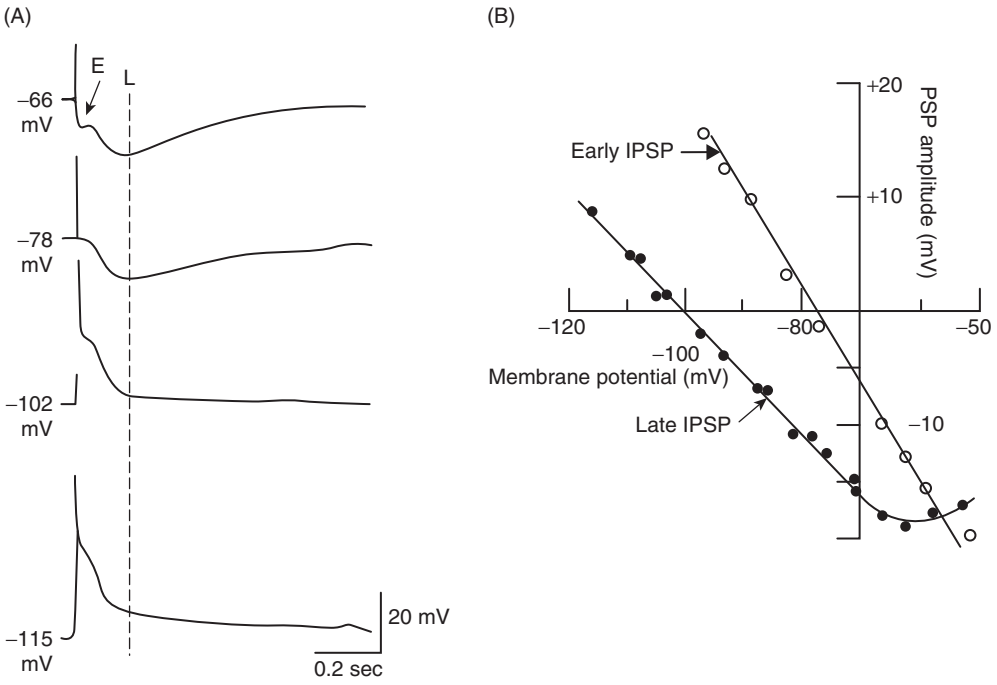


FIGURE 10.10 Dual-component IPSP. (A) Intracellular recordings from a pyramidal cell in the CA3 region of the rat hippocampus in response to activation of mossy fiber afferents. With the membrane potential of the cell at the resting potential, afferent stimulation produces an early (E) and late (L) IPSP. With increased hyperpolarization produced by injecting constant current into the cell, the early component reverses first. At more negative levels of the membrane potential, the late component also reverses. This result indicates that the ionic conductances underlying the two phases are distinct. (B) Plots of the change in amplitude of the early (measured at 25 ms) and the late (measured at 200 ms, dashed line in A) response as a function of membrane potential. Reversal potentials of the early and late components are consistent with a $GABA_A$-mediated chloride conductance and a $GABA_B$-mediated potassium conductance, respectively. *From Thalmann (1988).*

(Thalmann, 1988). The reversal potentials are consistent with a fast Cl^--mediated IPSP, mediated by fast opening of $GABA_A$ receptors and a slower K^+-mediated IPSP mediated by G-protein $GABA_B$ receptors.

Dual-component PSPs need not be strictly inhibitory or excitatory. For example, a presynaptic cholinergic neuron in the mollusk, *Aplysia*, produces a diphasic excitatory-inhibitory (E-I) response in its postsynaptic follower cell. The response can be simulated by local discrete application of ACh to the postsynaptic cell (Fig. 10.11) (Blankenship, Wachtel, & Kandel, 1971). The ionic mechanisms underlying this synaptic action were investigated in ion-substitution experiments, which revealed that the dual response is due to an early Na^+-dependent component followed by a slower Cl^--dependent component. Molecular mechanisms underlying such slow synaptic potentials are discussed next.

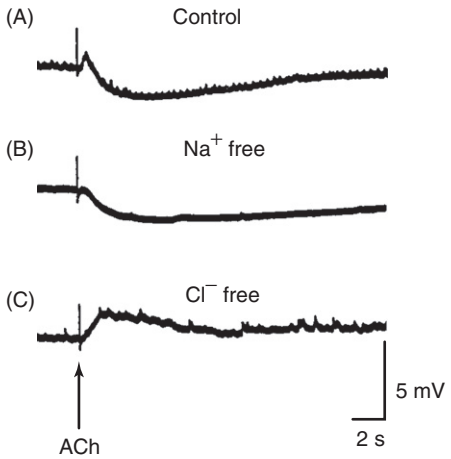


FIGURE 10.11 Dual-component cholinergic excitatory-inhibitory response. (A) Control in normal saline. Ejection of ACh produces a rapid depolarization followed by a slower hyperpolarization. (B) In Na^+-free saline, ACh produces a purely hyperpolarizing response, indicating that the depolarizing component in normal saline includes an increase in g_{Na}. (C) In Cl-free saline, ACh produces a purely depolarizing response, indicating that the hyperpolarizing component in normal saline includes an increase in g_{Cl}. *Reprinted with permission from Blankenship et al. (1971).*

Summary

Synaptic potentials mediated by ionotropic receptors are the fundamental means by which information is transmitted rapidly between neurons. Transmitters cause

channels to open in an all-or-none fashion, and the currents through these individual channels summate to produce the macroscopic postsynaptic potential. The sign of the postsynaptic potential is determined by the relationship between the membrane potential of the postsynaptic neuron and the ion selectivity of the ionotropic receptor.

METABOTROPIC RECEPTORS: MEDIATORS OF SLOW SYNAPTIC POTENTIALS

A common feature of the types of synaptic actions heretofore described is the direct binding of the transmitter with the receptor-channel complex. An entirely separate class of synaptic actions has as its basis the indirect coupling of the receptor with the channel. Two major types of coupling mechanisms have been identified: coupling of the receptor and channel through an intermediate regulatory protein, such as a G-protein, and coupling through a diffusible second-messenger system. Because coupling through a diffusible second-messenger system is the most common mechanism, it is the focus of this section.

A comparison of the features of direct, fast ionotropic-mediated and indirect, slow metabotropic-mediated synaptic potentials is shown in Figure 10.12. Slow synaptic potentials are not observed at every postsynaptic neuron, but Figure 10.12A illustrates an idealized case in which a postsynaptic neuron receives two inputs, one of which produces a conventional fast EPSP and the other of which produces a slow EPSP. An action potential in neuron 1 leads to an EPSP in the postsynaptic cell with a duration of about 30 ms (Fig. 10.12B). This type of potential might be produced in a spinal motor neuron by an action potential in an afferent fiber. Neuron 2 also produces a postsynaptic potential (Fig. 10.12C), but its duration (note the calibration bar) is more than three orders of magnitude greater than that of the EPSP produced by neuron 1.

How can a change in the postsynaptic potential of a neuron persist for many seconds as a result of a single action potential in the presynaptic neuron? Possibilities include a prolonged presence of the transmitter due to continuous release, to slow degradation, or to slow reuptake of the transmitter, but the mechanism here involves a transmitter-induced change in the metabolism of the postsynaptic cell. Figure 10.13 compares the general mechanisms for fast and slow synaptic potentials. Fast synaptic potentials are produced when a transmitter substance binds to a channel and produces a conformational change in the channel, causing it to become permeable to one or more ions (both Na^+ and K^+ in Fig. 10.13A). The increase in permeability leads

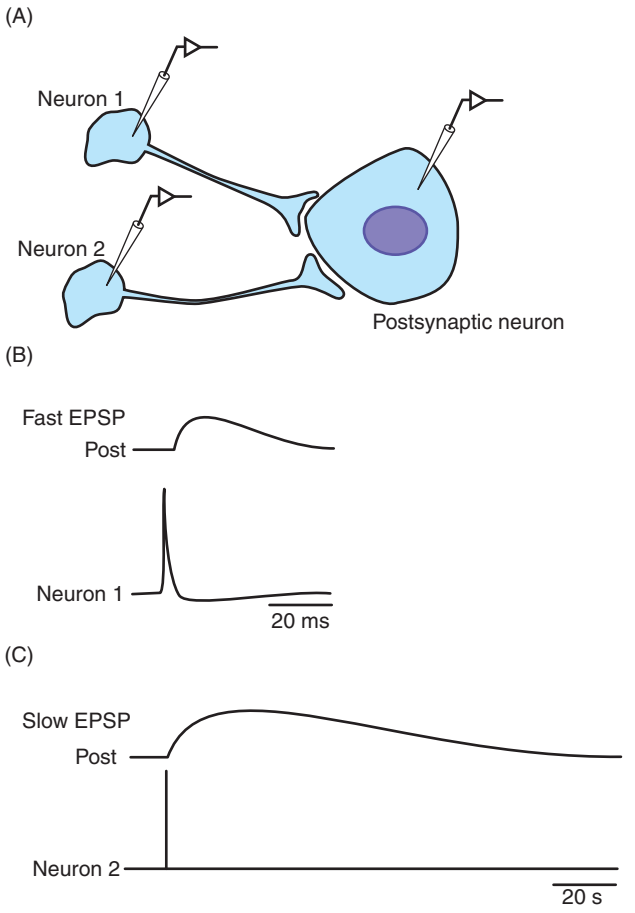


FIGURE 10.12 Fast and slow synaptic potentials. (A) Idealized experiment in which two neurons (1 and 2) make synaptic connections with a common postsynaptic follower cell (Post). (B) An action potential in neuron 1 leads to a conventional fast EPSP with a duration of about 30 ms. (C) An action potential in neuron 2 also produces an EPSP in the postsynaptic cell, but the duration of this slow EPSP is more than three orders of magnitude greater than that of the EPSP produced by neuron 1. Note the change in the calibration bar.

to a depolarization associated with the EPSP. The duration of the synaptic event critically depends on the amount of time during which the transmitter substance remains bound to the receptors. Acetylcholine, glutamate, and glycine remain bound only for a very short period. These transmitters are removed by diffusion, enzymatic breakdown, or reuptake into the presynaptic cell. Therefore, the duration of the synaptic potential is directly related to the lifetimes of the opened channels, and these lifetimes are relatively short (see Fig. 10.4B).

One mechanism for a slow synaptic potential is shown in Figure 10.13B. In contrast with the fast PSP for which the receptors are actually part of the ion channel complex, channels that produce slow synaptic potentials are not coupled directly to the transmitter receptors. Rather, the receptors are separated physically and exert their actions indirectly through changes

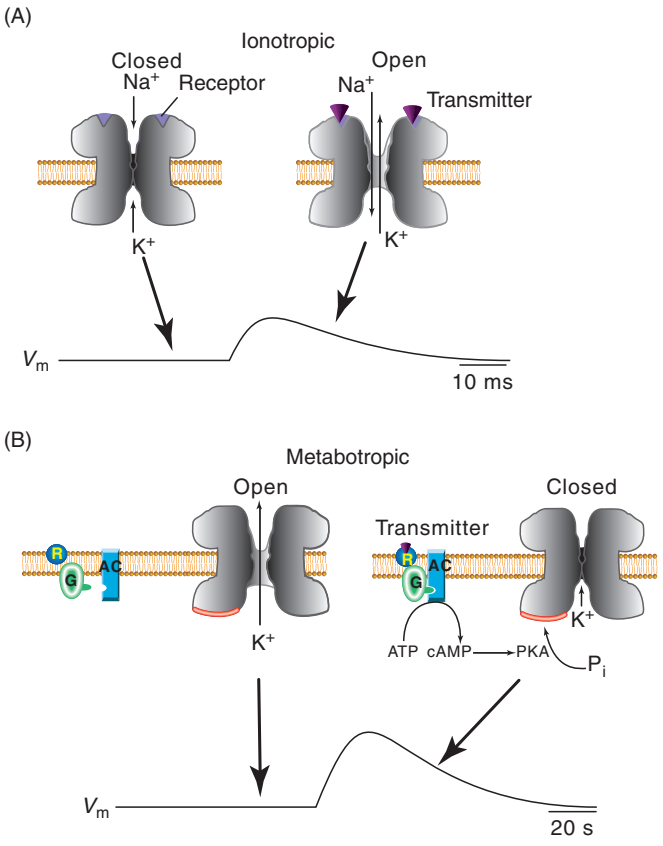


FIGURE 10.13 Ionotropic and metabotropic receptors and mechanisms of fast and slow EPSPs. (A, left) Fast EPSPs are produced by binding of the transmitter to specialized receptors that are directly associated with an ion channel (i.e., a ligand-gated channel). When the receptors are unbound, the channel is closed. (A, right) Binding of the transmitter to the receptor produces a conformational change in the channel protein such that the channel opens. In this example, the channel opening is associated with a selective increase in the permeability to Na^+ and K^+. The increase in permeability results in the EPSP shown in the trace. (B, left) Unlike fast EPSPs, which are due to the binding of a transmitter with a receptor-channel complex, slow EPSPs are due to the activation of receptors (metabotropic) that are not coupled directly to the channel. Rather, coupling takes place through the activation of one of several second-messenger cascades, in this example, the cAMP cascade. A channel that has a selective permeability to K^+ is normally open. (B, right) Binding of the transmitter to the receptor (R) leads to the activation of a G-protein (G) and adenylyl cyclase (AC). The synthesis of cAMP is increased, cAMP-dependent protein kinase (protein kinase A, PKA) is activated, and a channel protein is phosphorylated. The phosphorylation leads to closing of the channel and the subsequent depolarization associated with the slow EPSP shown in the trace. The response decays due to both the breakdown of cAMP by cAMP-dependent phosphodiesterase and the removal of phosphate from channel proteins by protein phosphatases (not shown).

in metabolism of specific second-messenger systems. Figure 10.13B illustrates one type of response in *Aplysia* for which the cAMP-protein kinase A (PKA) system is the mediator, but other slow PSPs use other second-messenger kinase systems (e.g., the protein kinase C system). In the cAMP-dependent slow synaptic responses in *Aplysia*, transmitter binding to membrane receptors activates G-proteins and stimulates an increase in the synthesis of cAMP. Cyclic AMP then leads to the activation of cAMP-dependent protein kinase (PKA), which phosphorylates a channel protein or protein associated with the channel (Siegelbaum, Camardo, & Kandel, 1982). A conformational change in the channel is produced, leading to a change in ionic conductance. Thus, in contrast with a direct conformational change produced by the binding of a transmitter to the receptor-channel complex, in this case, a conformational change is produced by protein phosphorylation. Indeed, phosphorylation-dependent channel regulation is a fairly general feature of slow PSPs. However, channel regulation by second messengers is not produced exclusively by phosphorylation. In one family of ion channels, the channels are gated or regulated directly by cyclic nucleotides. These cyclic nucleotide-gated channels require cAMP or cGMP to open but have other features in common with members of the superfamily of voltage-gated ion channels (Kaupp, 1995; Zimmermann, 1995).

Another interesting feature of slow synaptic responses is that they are sometimes associated with decreases rather than increases in membrane conductance. For example, the particular channel illustrated in Figure 10.13B is selectively permeable to K^+ and is normally open. As a result of the activation of the second messenger, the channel closes and becomes less permeable to K^+. The resultant depolarization may seem paradoxical, but recall that the membrane potential is due to a balance between resting K^+ and Na^+ permeability. K^+ permeability tends to move the membrane potential toward the K^+ equilibrium potential (−80 mV), whereas Na^+ permeability tends to move the membrane potential toward the Na^+ equilibrium potential (55 mV). Normally, K^+ permeability predominates, and the resting membrane potential is close to, but not equal to, the K^+ equilibrium potential. If K^+ permeability is decreased because some of the channels close, the membrane potential will be biased toward the Na^+ equilibrium potential and the cell will depolarize.

At least one reason for the long duration of slow PSPs is that second-messenger systems are slow (from seconds to minutes). Take the cAMP cascade as an example. Cyclic AMP takes some time to be synthesized, but, more importantly, after synthesis, cAMP levels can remain elevated for a relatively long period (minutes). The duration of the elevation of cAMP depends on the actions of cAMP-phosphodiesterase, which breaks down cAMP. However, duration of an effect could outlast the duration of the change in the second messenger because of persistent phosphorylation of the substrate protein(s). Phosphate groups are removed from substrate proteins by protein phosphatases. Thus, the net duration of a

response initiated by a metabotropic receptor depends on the actions of not only the synthetic and phosphorylation processes, but also the degradative and dephosphorylation processes.

Activation of a second messenger by a transmitter can have a localized effect on the membrane potential through phosphorylation of membrane channels near the site of a metabotropic receptor. The effects can be more widespread and even longer lasting than depicted in Figure 10.13B. For example, second messengers and protein kinases can diffuse and affect more distant membrane channels. Moreover, a long-term effect can be induced in the cell by altering gene expression. For example, protein kinase A can diffuse to the nucleus, where it can activate proteins that regulate gene expression. Detailed descriptions of second messengers and their actions are given in Chapter 9.

Summary

In contrast to the rapid responses mediated by ionotropic receptors, responses mediated by metabotropic receptors are generally relatively slow to develop and persistent. These properties arise because metabotropic responses can involve the activation of second-messenger systems. By producing slow changes in the resting potential, metabotropic receptors provide long-term modulation of the effectiveness of responses generated by ionotropic receptors. Moreover, these receptors, through the engagement of second-messenger systems, provide a vehicle by which a presynaptic cell can not only alter the membrane potential, but also produce widespread changes in the biochemical state of a postsynaptic cell.

INTEGRATION OF SYNAPTIC POTENTIALS

The small amplitude of the EPSP in spinal motor neurons (and other cells in the CNS) poses an interesting question. Specifically, how can an EPSP with an amplitude of only 1 mV drive the membrane potential of the motor neuron (i.e., the postsynaptic neuron) to threshold and fire the spike in the motor neuron that is necessary to produce the contraction of the muscle? The answer to this question lies in the principles of temporal and spatial summation.

When the ligament is stretched (Fig. 10.1), many stretch receptors are activated. Indeed, the greater the stretch, the greater the probability of activating a larger number of the stretch receptors; this process is referred to as recruitment. However, recruitment is not the complete story. The principle of frequency coding in the nervous system specifies that the greater the intensity of a stimulus, the greater the number of action potentials per unit time (frequency) elicited in a sensory neuron. This principle applies to stretch receptors as well. Thus, the greater the stretch, the greater the number of action potentials elicited in the stretch receptor in a given interval and therefore the greater the number of EPSPs produced in the motor neuron from that train of action potentials in the sensory cell. Consequently, the effects of activating multiple stretch receptors add together (spatial summation), as do the effects of multiple EPSPs elicited by activation of a single stretch receptor (temporal summation). Both of these processes act in concert to depolarize the motor neuron sufficiently to elicit one or more action potentials, which then propagate to the periphery and produce the reflex.

Temporal Summation Allows Integration of Successive PSPs

Temporal summation can be illustrated by firing action potentials in a presynaptic neuron and monitoring the resultant EPSPs. For example, in Figures 10.14A and 10.14B, a single action potential in sensory neuron 1 produces a 1-mV EPSP in the motor neuron. Two action potentials in quick succession produce two EPSPs, but note that the second EPSP occurs during the falling phase of the first, and the depolarization associated with the second EPSP adds to the depolarization produced by the first. Thus, two action potentials produce a summated potential that is about 2 mV in amplitude. Three action potentials in quick succession would produce a summated potential of about 3 mV. In principle, 30 action potentials in quick succession would produce a potential of about 30 mV and easily drive the cell to threshold. This summation is strictly a passive property of the cell. No special ionic conductance mechanisms are necessary. Specifically, the postsynaptic conductance change (g_{syn} in Eq. 10.3) produced by the second of two successive action potentials adds to that produced by the first. In addition, the postsynaptic membrane has a capacitance and can store charge. Thus, the membrane temporarily stores the charge of the first EPSP, and the charge from the second EPSP is added to that of the first.

However, the "time window" for this process of temporal summation very much depends on the duration of the postsynaptic potential, and temporal summation is possible only if the presynaptic action potentials (and hence postsynaptic potentials) are close in time to each other. The time frame depends on the duration of changes in the synaptic conductance and the time constant (Chapter 5). Temporal summation, however, rarely is observed to be linear as in the preceding examples, even when the postsynaptic conductance change (g_{syn} in Eq. 10.3) produced by the second of two successive action potentials is identical with that produced by the

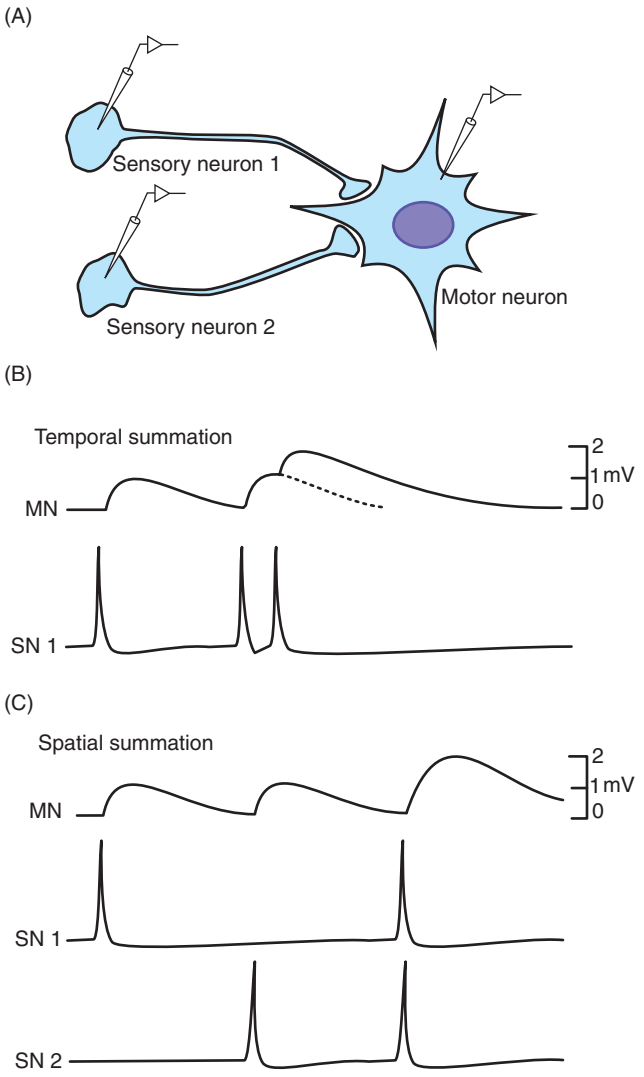


FIGURE 10.14 Temporal and spatial summation. (A) Intracellular recordings are made from two idealized sensory neurons (SN1 and SN2) and a motor neuron (MN). (B) Temporal summation. A single action potential in SN1 produces a 1-mV EPSP in the MN. Two action potentials in quick succession produce a dual-component EPSP, the amplitude of which is approximately 2 mV. (C) Spatial summation. Alternative firing of single action potentials in SN1 and SN2 produce 1-mV EPSPs in the MN. Simultaneous action potentials in SN1 and SN2 produce a summated EPSP, the amplitude of which is about 2 mV.

first (i.e., no presynaptic facilitation or depression) and the synaptic current is slightly less because the first PSP reduces the driving force ($V_m - E_r$) for the second. Interested readers should try some numerical examples.

Spatial Summation Allows Integration of PSPs from Different Parts of a Neuron

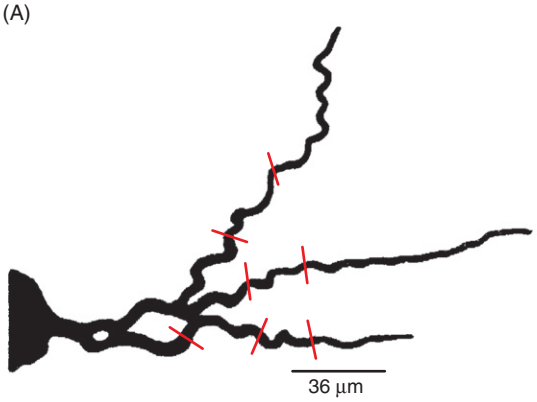
Spatial summation (Fig. 10.14C) requires a consideration of more than one input to a postsynaptic neuron. An action potential in sensory neuron 1 produces a 1-mV EPSP, just as it did in Figure 10.14B. Similarly, an action potential in a second sensory neuron by itself also produces a 1-mV EPSP. Now, consider the consequences of action potentials elicited simultaneously in sensory neurons 1 and 2. The net EPSP is equal to the summation of the amplitudes of the individual EPSPs. Here, the EPSP from sensory neuron 1 is 1 mV, the EPSP from sensory neuron 2 is 1 mV, and the summated EPSP is approximately 2 mV (Fig. 10.14C). Thus, spatial summation is a mechanism by which synaptic potentials generated at different sites can summate. Spatial summation in nerve cells is influenced by the space constant—the ability of a potential change produced in one region of a cell to spread passively to other regions of a cell (see Chapter 11).

SYNAPTIC PLASTICITY

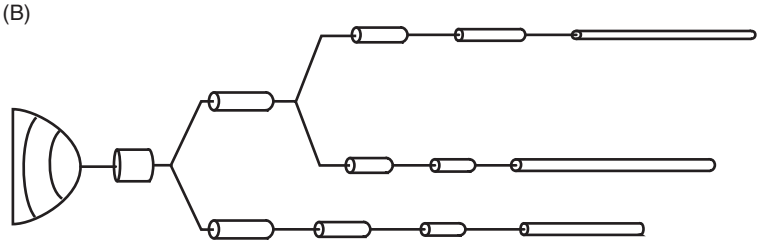
Synaptic plasticity refers to the ability of synaptic connections to change their strength, a process that is believed critical for regulating the flow of information in neuronal circuits and for mediating learning and memory. Synaptic plasticity has two major forms, homosynaptic plasticity and heterosynaptic plasticity. Homosynaptic plasticity refers to the ability of a synapse to change its strength as a result of activity in that synapse, whereas heterosynaptic plasticity refers to the ability of a synapse to change its strength as a result of activity in another neuron or pathway.

Homosynaptic plasticity has two basic forms: homosynaptic depression and homosynaptic facilitation. Homosynaptic depression refers to a decrease in the strength of a synaptic connection with repeated activation. For example, if the afferent sensory neuron in Figure 10.2 is stimulated a second time 100 ms after the first, the second EPSP will have a lower amplitude. The mechanisms underlying synaptic depression are diverse and range from depletion of the pools of transmitter in the presynaptic terminal to desensitization of transmitter receptors on the postsynaptic neuron. The various possibilities are not mutually exclusive. Homosynaptic facilitation is a phenomenon opposite to homosynaptic depression. It generally occurs when multiple action potentials are fired in quick succession in a presynaptic neuron. For example, if the afferent sensory neuron in Figure 10.2 is stimulated a second time 10 ms after the first, the second EPSP will have a greater amplitude. This particular example of homosynaptic facilitation is frequently called paired-pulse or twin-pulse facilitation (PPF). A common mechanism for PPF involves residual calcium. If the second presynaptic action potential is initiated during a time when Ca^{2+} levels remain elevated after the first, the Ca^{2+} influx from the second spike will add to it. Consequently, more transmitter will be released because of the greater net Ca^{2+} concentration in the

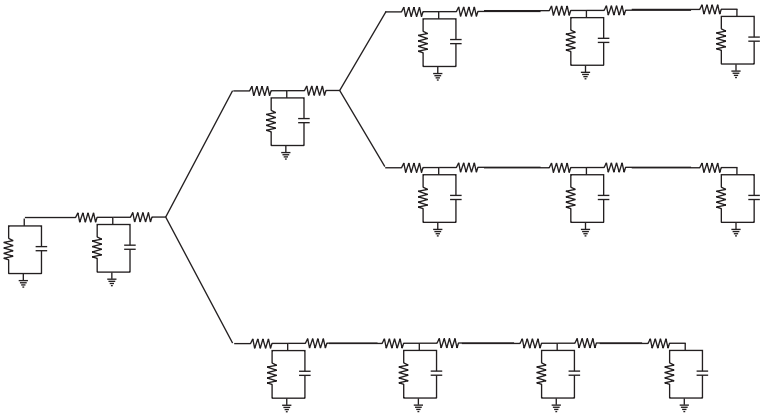
(A)


36 μm

(B)



(C)



(D)

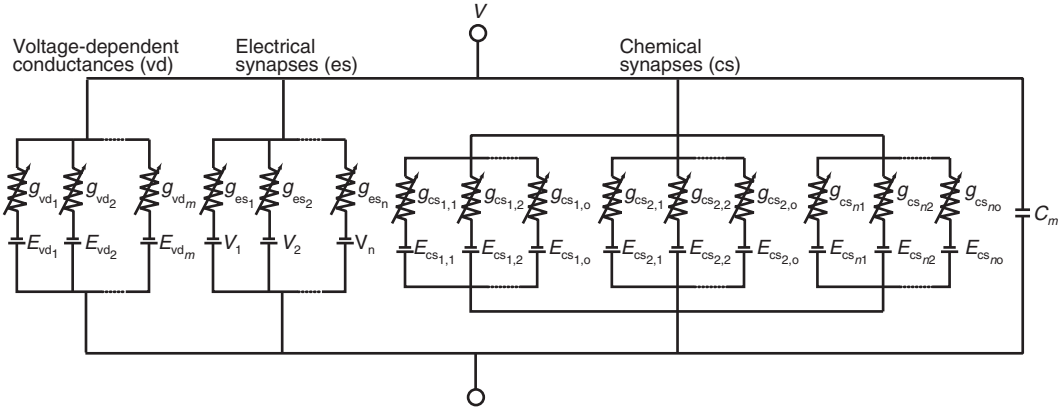

V
Voltage-dependent conductances (vd)
Electrical synapses (es)
Chemical synapses (cs)
g_{vd_1} g_{vd_2} g_{vd_m}
E_{vd_1} E_{vd_2} E_{vd_m}
g_{es_1} g_{es_2} g_{es_n}
V_1 V_2 V_n
$g_{cs_{1,1}}$ $g_{cs_{1,2}}$ $g_{cs_{1,o}}$
$E_{cs_{1,1}}$ $E_{cs_{1,2}}$ $E_{cs_{1,o}}$
$g_{cs_{2,1}}$ $g_{cs_{2,2}}$ $g_{cs_{2,o}}$
$E_{cs_{2,1}}$ $E_{cs_{2,2}}$ $E_{cs_{2,o}}$
$g_{cs_{n1}}$ $g_{cs_{n2}}$ $g_{cs_{no}}$
$E_{cs_{n1}}$ $E_{cs_{n2}}$ $E_{cs_{no}}$
C_m

FIGURE 10.15 Modeling the integrative properties of a neuron. (A) Partial geometry of a neuron in the CNS revealing the cell body and pattern of dendritic branching. (B) The neuron modeled as a sphere connected to a series of cylinders, each of which represents the specific electrical properties of a dendritic segment. (C) Segments linked with resistors representing the intracellular resistance between segments, with each segment represented by the parallel combination of the membrane capacitance and the total membrane conductance. *Reprinted with permission from Koch and Segev (1989). Copyright © 1989 MIT Press.* (D) Electrical circuit equivalent of the membrane of a segment of a neuron. The segment has a membrane potential V and a membrane capacitance C_m. Currents arise from three sources: (1) m voltage-dependent (vd) conductances (g_{vd1}–g_{vdm})/(2) n conductances due to electrical synapses (es) (g_{es1}–g_{esn}), and (3) n times o time-dependent conductances due to chemical synapses (cs) with each of the n presynaptic neurons ($g_{cs1,1}$–$g_{csn,o}$). E_{vd} and E_{cs} are constants and represent the values of the equilibrium potential for currents due to voltage-dependent conductances and chemical synapses, respectively. V_1–V_n represent the value of the membrane potential of the coupled cells. *Reprinted with permission from Ziv et al. (1994).*

presynaptic terminal. When two spikes are close together PPF occurs rather than homosynaptic depression presumably because the increase in the probably of release caused by the enhanced Ca^{2+} is much greater than the decrease in the pool size due to the depletion of transmitter. A more dramatic example of homosynaptic facilitation is post-tetanic potentiation (PTP). Here, rather than two spikes a high frequency burst of spikes (tetanus) is delivered to the presynaptic neuron and subsequent test EPSPs are enhanced.

Like homosynaptic plasticity, heterosynaptic plasticity has two forms: heterosynaptic inhibition and heterosynaptic facilitation. In both cases, the strength of the synapse between two neurons is regulated by a third neuron or pathway. The prime example of heterosynaptic inhibition is presynaptic inhibition in the spinal cord where descending control pathways regulate the strength of connections between sensory neurons and their postsynaptic targets. At the mechanistic level, presynaptic inhibition appears to be due to a down regulation of voltage-dependent Ca^{2+} conductances in the presynaptic terminals. Consequently, action potentials in the sensory neurons lead to less influx of Ca^{2+} and less transmitter release. Therefore, the amplitude of the EPSP in the postsynaptic cell is reduced (i.e., inhibited). The most extensively studied example of heterosynaptic facilitation is the serotonin-induced facilitation of the *Aplysia* sensorimotor synapse, which has both short lasting (min) and long lasting (days) forms. This process is described in detail in Chapter 47. Other more enduring forms of synaptic plasticity such as long-term potentiation (LTP) and long-term depression (LTD) are also described in Chapter 47.

Summary

Whether a neuron fires in response to synaptic input depends, at least in part, on how many action potentials are produced in any one presynaptic excitatory pathway and on how many individual convergent excitatory input pathways are activated. The summation of EPSPs in time and space is only part of the process, however. The final behavior of the cell is also due to the summation of inhibitory synaptic inputs in time and space, as well as to the properties of the voltage-dependent currents (Fig. 10.15) in the soma and along the dendrites (Koch & Segev, 1989; Ziv, Baxter, & Byrne, 1994). For example, voltage-dependent conductances such as A-type K^+ conductance have a low threshold for activation and can thus oppose the effectiveness of an EPSP to trigger a spike. Low-threshold Na^+ and Ca^{2+} channels can boost an EPSP. Finally, we need to consider that spatial distribution of the various voltage-dependent channels, ligand-gated receptors, and metabotropic receptors is not uniform. Thus, each segment of the neuronal membrane can perform selective integrative functions. Clearly, this system has an enormous capacity for the local processing of information and for performing logical operations.

The flow of information in dendrites and the local processing of neuronal signals are discussed in Chapter 12. Several software packages are available for the development and simulation of realistic models of single neurons and neural networks. One, Simulator for Neural Networks and Action Potentials (SNNAP) (http://nba.uth.tmc.edu/snnap/), provides mathematical descriptions of ion currents, intracellular second messengers, and ion pools, and allows simulation of current flow in multicompartment models of neurons.

References

Billinton, A., Ige, A. O., Bolam, J. P., White, J. H., Marshall, F. H., & Emson, P. C. (2001). Advances in the molecular understanding of $GABA_B$ receptors. *Trends in Neurosciences, 24,* 277–282.

Blankenship, J. E., Wachtel, H., & Kandel, E. R. (1971). Ionic mechanisms of excitatory, inhibitory and dual synaptic actions mediated by an identified interneuron in abdominal ganglion of *Aplysia*. *Journal of Neurophysiology, 34,* 76–92.

Bormann, J. (1988). Electrophysiology of $GABA_A$ and $GABA_B$ receptor subtypes. *Trends in Neurosciences, 11,* 112–116.

Bormann, J., & Feigenspan, A. (1995). $GABA_C$ receptors. *Trends in Neurosciences, 18,* 515–519.

Bowery, N. G. (1993). $GABA_B$ receptor pharmacology. *Annual Review of Pharmacology and Toxicology, 33,* 109–147.

Cherubini, E., & Conti, F. (2001). Generating diversity at GABAergic synapses. *Trends in Neurosciences, 24,* 155–162.

Eccles, J. C. (1964). *The physiology of synapses*. New York: Springer-Verlag.

Fatt, P., & Katz, B. (1951). An analysis of the end-plate potential recorded with an intra cellular electrode. *Journal of Physiology (London), 115,* 320–370.

Gage, P. W. (1992). Activation and modulation of neuronal K^+ channels by GABA. *Trends in Neurosciences, 15,* 46–51.

Hamill, O. P., Marty, A., Neher, E., Sakmann, B., & Sigworth, J. (1981). Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pflügers Archives, 391,* 85–100.

Hestrin, S., Nicoll, R. A., Perkel, D. J., & Sah, P. (1990). Analysis of excitatory synaptic action in pyramidal cells using whole-cell recording from rat hippocampal slices. *Journal of Physiology (London), 422,* 203–225.

Johnston, D., & Wu, S. M. -S. (1995). *Foundations of cellular neurophysiology*. Cambridge, MA: MIT Press.

Kaupp, U. B. (1995). Family of cyclic nucleotide gated ion channels. *Current Opinion Neurobiology, 5,* 434–442.

Koch, C., & Segev, I. (1989). *Methods in neuronal modeling*. Cambridge, MA: MIT Press.

Moss, S. J., & Smart, T. G. (2001). Constructing inhibitory synapses. *Nature Reviews Neuroscience, 2,* 240–250.

Sakmann, B. (1992). Elementary steps in synaptic transmission revealed by currents through single ion channels. *Science, 256,* 503–512.

Siegelbaum, S. A., Camardo, J. S., & Kandel, E. R. (1982). Serotonin and cyclic AMP close single K^+ channels in *Aplysia* sensory neurones. *Nature (London), 299,* 413–417.

Spencer, W. A. (1977). The physiology of supraspinal neurons in mammals. In E. R. Kandel (Ed.), *Handbook of physiology* (Part 2, Sect. 1, Vol. 1, pp. 969–1022). Bethesda, MD: American Physiological Society.

Takeuchi, A., & Takeuchi, N. (1960). On the permeability of end-plate membrane during the action of transmitter. *Journal of Physiology (London), 154,* 52–67.

Thalmann, R. H. (1988). Evidence that guanosine triphosphate (GTP)-binding proteins control a synaptic response in brain: Effect of pertussis toxin and GTPgS on the late inhibitory post-synaptic potential of hippocampal CA3 neurons. *Journal of Neuroscience, 8,* 4589–4602.

Zimmermann, A. L. (1995). Cyclic nucleotide gated channels. *Current Opinion Neurobiology, 5,* 296–303.

Ziv, I., Baxter, D. A., & Byrne, J. H. (1994). Simulator for neural networks and action potentials: Description and application. *Journal of Neurophysiology, 71,* 294–308.

Suggested Readings

Burke, R. E., & Rudomin, P. (1977). Spatial neurons and synapses. In E. R. Kandel (Ed.), *Handbook of physiology* (Part 2, Sect. 1, Vol. 1, pp. 877–944). Bethesda, MD: American Physiological Society.

Byrne, J. H., & Schultz, S. G. (1994). *An introduction to membrane transport and bioelectricity* (2nd ed.). New York: Raven Press.

Cowan, W. M., Sudhof, T. C., & Stevens, C. F. (Eds.). (2001). *Synapses*. Johns Hopkins Univ. Press.

Hille, B. (Ed.). (2001). *Ion channels of excitable membranes* (3rd ed.). Sunderland, MA: Sinauer.

Shepherd, G. M. (Ed.). (2004). *The synaptic organization of the brain* (5th ed.). New York: Oxford Univ. Press.

John H. Byrne

CHAPTER

11

Information Processing in Dendrites and Spines

INTRODUCTION

A central goal of neuroscience is to understand how neural circuits control sensation and action, perception, and cognition. These circuits are constructed of neurons, most of which have elaborately branching dendritic trees. Dendrites receive and integrate a wealth of information; in many cases tens of thousands of synapses are formed on the dendritic tree of a single neuron. Activation of these synapses results in complex spatiotemporal signal integration involving current flow that ultimately converges in the axon, where "decisions" are made regarding action potential initiation. Signal integration in dendrites is central to the function of neural circuits and therefore contributes to the brain's ability to perform the multimodal integrative tasks that are key to survival. Thus, a detailed analysis of how synaptic inputs are integrated in dendrites is essential to an understanding of the neural control of behavior and cognition.

A traditional approach to investigating the function of neural circuits is through modeling of neural networks. Typically, these models are built using simple neurons that integrate excitatory and inhibitory inputs until a fixed threshold for action potential firing is reached. Some real neurons might function like this. For example, the most numerous cell type in the brain—the granule cells of the cerebellum—integrates just a few excitatory inputs on very small dendrites. In these neurons, synaptic integration is relatively simple. Most other neurons in the brain, however, have much more elaborate dendritic trees, and accordingly, synaptic integration is much more complex.

The structure of dendritic trees is very diverse (Fig. 11.1), likely reflecting diversity in the functional properties and the types of computations performed by different types of neurons. To fully understand synaptic integration and, ultimately, the computations performed by neurons and circuits, it is necessary to achieve an appreciation of how synaptic inputs are processed in a variety of neuronal types. Dendritic integration has been studied in several types of neurons, with an emphasis on pyramidal neurons of the neocortex and hippocampus and Purkinje cells of the cerebellum. From this work a number of general principles have been learned, which will facilitate the analysis of dendritic integration in other cell types as the field advances.

How does the presence of dendrites affect synaptic integration? The answer is complex, but it has its roots in the fact that the interactions between multiple synapses, and the mechanisms that ultimately lead to action potential firing in the axon, are determined by current flow both across the membrane and longitudinally along dendritic branches. Thus, a key component to understanding how synaptic inputs are integrated is to quantitatively analyze how current flows in dendritic trees. This first became possible with the development of neuronal cable theory by Wilfrid Rall (Box 11.1).

Rall took the mathematics originally developed to describe the flow of current (and the resultant voltage changes) in transatlantic telegraph cables and applied it to branching structures, in particular dendrites. Rall's theory is sometimes called "passive cable theory," because it is based on the passive membrane properties of neurons (see Chapter 6). Rall recognized, however, that not all neuronal cables are passive; axons certainly are not, and at the time the theory was developed, it was already becoming clear that dendrites might also be excitable (i.e., contain voltage-gated channels that provide inward current). Rall nevertheless realized that

© 2013 Elsevier Inc. All rights reserved.

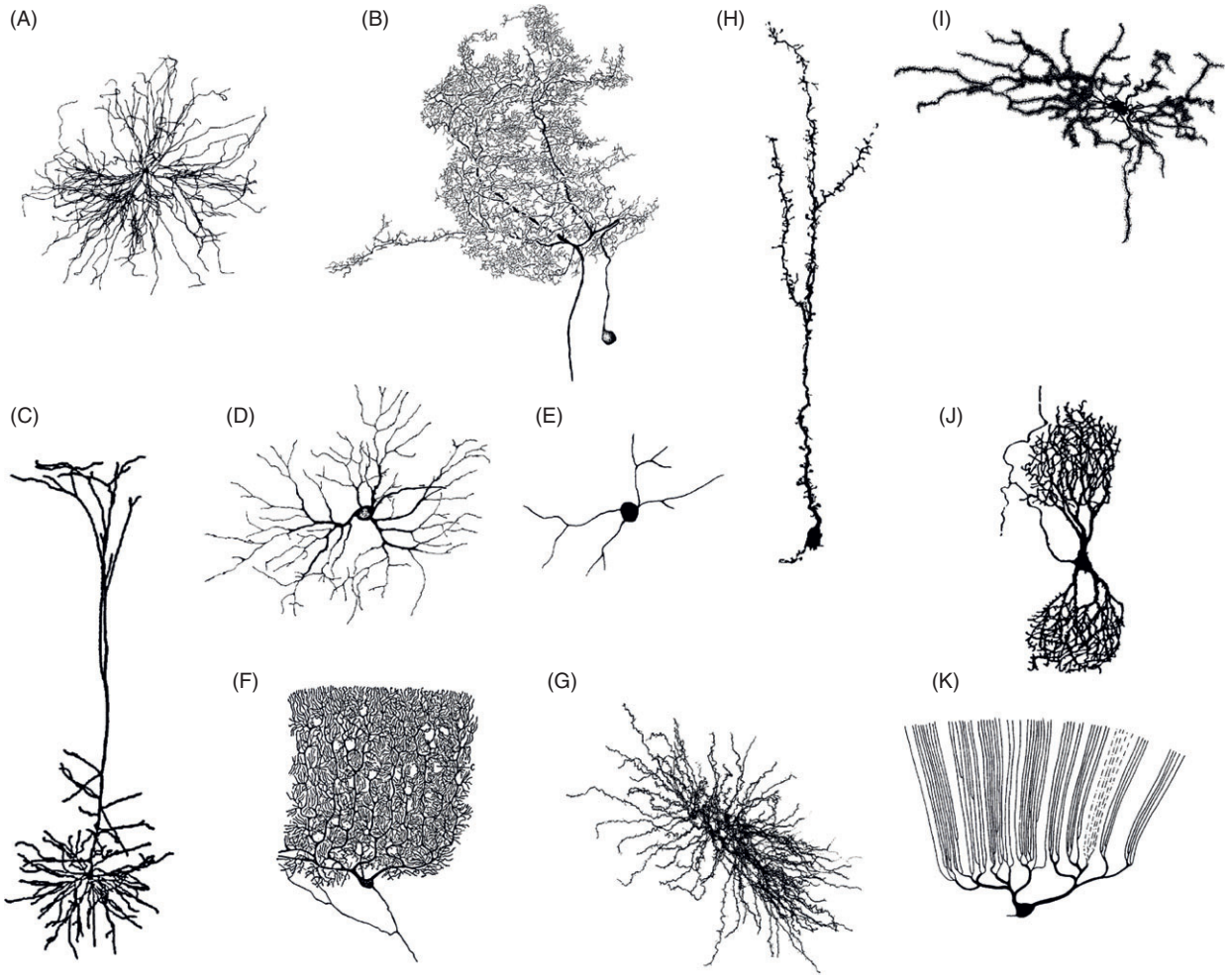


FIGURE 11.1 Variation in the structure of dendritic trees. Reconstructions are shown from: (A) alpha motoneuron of cat spinal cord, (B) spiking interneuron from mesothoracic ganglion of locust, (C) neocortical layer 5 pyramidal neuron of rat, (D) retinal ganglion cell in cat, (E) amacrine cell from retina of larval salamander, (F) cerebellar Purkinje cell in human, (G) relay neuron in basoventral thalamus of rat, (H) granule cell from olfactory bulb of mouse, (I) spiny projection neuron from striatum of rat, (J) nerve cell in nucleus of Burdach of human fetus, (K) Purkinje cell of mormyrid fish. *Source: Mel, 1994.*

BOX 11.1

NEURONAL CABLE THEORY AND COMPUTATIONAL MODELING

One of the seminal contributions to neuroscience is neuronal cable theory, which was developed by Wilfrid Rall in the 1950s–1960s. The theory provided a framework for quantitatively analyzing the effects of dendrites and dendritic spines on the process of synaptic integration. The central concept is that dendrites are conductive cables separated from a conducting surrounding medium (the extracellular space) by a mostly insulating membrane. The cable is characterized by its intracellular resistivity (R_i), its membrane resistivity (R_m), and its specific membrane capacitance (C_m). If the membrane potential is changed at the sealed end of a cable that extends infinitely in one direction, the voltage will decay exponentially, according to the equation:

$$\Delta V_x = \Delta V_0 * \exp(-x/\lambda)$$

where ΔV_0 is the voltage change imposed at the end, ΔV_x is the voltage change at distance x along the cable, and λ is the "space constant" of the cable, given by $\lambda = [(d/4)(R_m/R_i)]^{1/2}$, where d is the diameter of the cable. For transient voltage changes, the terminal decay of the voltage over time is governed by the membrane time constant, $\tau = R_m C_m$.

BOX 11.1 (*cont'd*)

The mathematical centerpiece of neuronal cable theory is the cable equation:

$$\frac{\partial V}{\partial T} = \frac{\partial^2 V}{\partial X^2} - V$$

This differential equation describes the change in voltage along a cable in both time and space, where X and T are dimensionless variables ($X = x/\lambda$ and $T = t/\tau$). The cable equation had been used previously for axons (Cole & Hodgkin, 1939; Davis & Lorente de Nó, 1947; Hodgkin & Rushton, 1946), but Rall extended its application to dendrites. Although much of Rall's work used this equation to analyze voltage changes in simple linear cables, he also applied it to branching cables and showed that it could be used to analyze dendrites with arbitrary branching geometries. Indeed, one of Rall's (1959) key contributions was his analysis of the effects of branching in cables: primarily in dendrites, but also in axons (Goldstein & Rall, 1974).

In its analytical form, the cable equation can only be applied to passive dendrites with current sources (i.e., no synaptic or voltage-gated conductances). Rall realized, however, that dendrites were not likely to function as purely passive cables, so he also applied the cable equation to numerical simulations in order to facilitate simulation of more complex conditions, such as the activation of time- and voltage-dependent conductances in dendrites. Central to this approach is the concept that each segment of a dendrite can be broken up into a number of isopotential compartments, with changes in membrane potential between compartments computed using the cable equation with discrete time steps. The accuracy of such simulations therefore depends on the assignment of appropriately small compartments and short time steps.

Rall's calculations and computer simulations provided myriad insights, including many of the principles described in this chapter. Two of them are illustrated here: the distance-dependent filtering of synaptic potentials along cables and the difference in responses to groups of synapses activated in different spatio-temporal patterns along a dendrite (Fig. B11.1). His many other findings are reviewed in the book *Theoretical Foundation of Dendritic Function* (Segev, Rinzel, & Shepherd, 1995), which compiles and summarizes Rall's most important contributions to the field.

Two other aspects of Rall's work are just as important as his papers. First, Rall pioneered the use of computers to simulate not only single neurons but also small circuits of neurons. Second, Rall realized the importance of a close alliance between theorists and experimentalists, and his work is a testament to the progress that can be achieved using this approach.

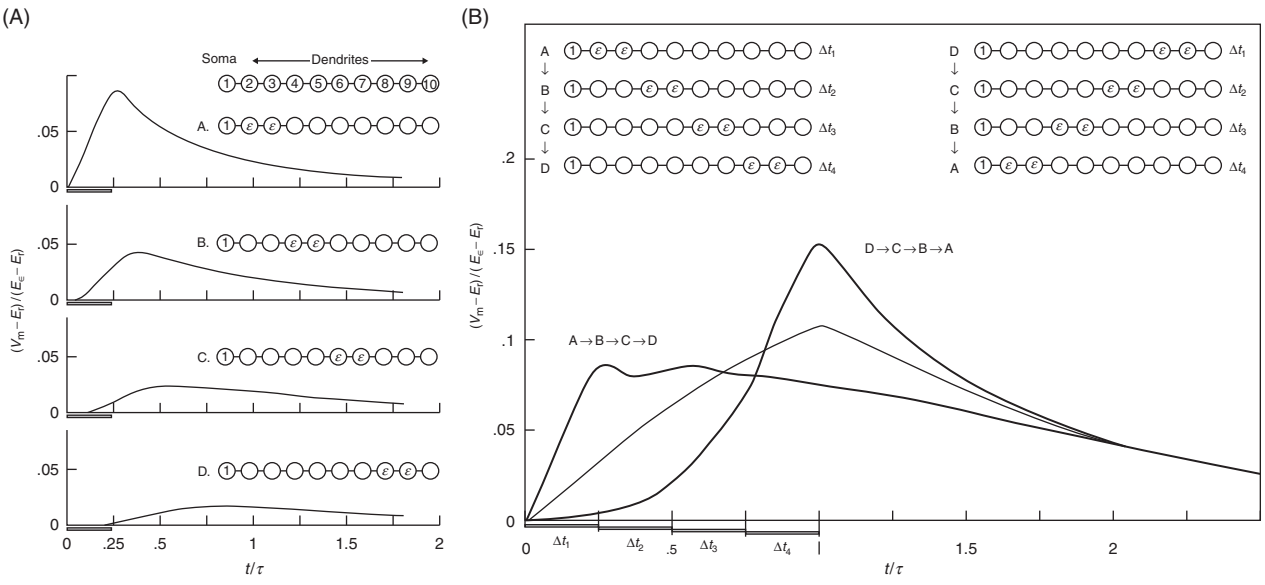


FIGURE B11.1 Simulations of excitatory synaptic inputs in a model dendrite. (A) Somatic membrane potential for excitatory synapses (ε) positioned in different compartments of a dendrite. Synapses positioned further from the soma resulted in progressively smaller and slower somatically recorded synaptic potentials. EPSP amplitude is normalized to the driving force and time is normalized to the membrane time constant. (B) Somatic membrane potential for activation of excitatory synapses in opposite order. For A->B->C->D the proximal synapses were activated first; for D->C->B->A the distal synapses were activated first. Synapses were simulated at temporal increments of 0.25τ. *Source: Rall (1964).*

passive cable theory could provide the necessary foundation for a more in-depth analysis of the complex processing that would occur in excitable dendrites. He also showed that cable theory could be used to develop computational models of elaborately branching dendritic trees. Computational modeling based on cable theory has taken a central place in the field of neuroscience, allowing for complex integrative processes to be analyzed quantitatively, leading to new insights and experiments that have resulted in new discoveries. In fact, in the decades that have followed Hodgkin and Huxley's landmark experimental and theoretical analysis of the action potential and its propagation along axons, the field of dendritic integration is one of the areas in neuroscience where the combination of experimental and theoretical approaches has been employed very effectively.

Following the logical sequence pioneered by Rall, we begin by considering how the integration of synaptic inputs occurs in passive dendrites, and continue by considering how synaptic integration is affected by active dendritic properties.

SYNAPTIC INTEGRATION IN PASSIVE DENDRITES

Decades ago, Rall used his theory to predict that somatically recorded synaptic potentials would be much smaller when the activated synapses are located in distal dendrites than if the same synapses were located close to the soma. In addition, he predicted that the filtering effects of dendrites would have effects on the time course of EPSPs and the interactions between synapses. More recently, Rall's predictions have been validated by experimental measurements, made possible by the development of methods for recording directly from the dendrites of neurons (Box 11.2).

Effects of Dendrites on Synaptic Potentials at the Soma

In most neurons, the soma is an important location for a number of reasons. Dendrites converge on the soma, forming a confluence for current flowing from the dendrites (Fig. 11.2A). Current that reaches the soma flows from there to the axon, which usually emerges at or very near the soma. This integrated current changes the membrane potential in the axon initial segment, where action potentials are initiated if the depolarization is sufficiently large. The soma also happens to be the location at which most electrophysiological recordings are obtained. This is convenient, because somatic recordings reflect the integration occurring at a location very near the action potential initiation zone, but it is important to understand how the presence of dendrites influences the current flowing into and out of the soma and how this shapes the membrane potential that is recorded there. The magnitude of the dendritic filtering of synaptic potentials depends on the overall structure of the dendritic tree, the membrane properties of dendrites, and the location of the synapses. In most neurons, dendrites exert a powerful filtering effect on synaptic potentials spreading through them, and as a result somatically recorded potentials are much smaller than the same events recorded in the dendrites (Fig. 11.2).

Two effects influence the difference between somatic and dendritically recorded EPSPs. First, some of the charge that enters the neuron at a synapse never reaches the soma. When a synapse depolarizes a dendrite, there is a voltage difference between the dendrite and the soma, which will promote current flow toward the soma (where the direction of current flow is defined by the movement of positive charge). Current travels within dendrites via electron transfer between charged particles in the cytoplasm, but because the density of charged particles is relatively low (compared to a copper wire, for instance), dendrites have a significant longitudinal resistance. As a result, some current will flow along alternate paths, such as onto the capacitance of the dendritic membrane and across its resistance (Fig. 11.2B). Most dendritic trees have a lot of branches and a large surface area, so there can be significant loss of synaptic current as it flows into dendritic branches on its way to the soma. In fact, in many cases only a small fraction of the total synaptic current reaches the soma and contributes to the local depolarization observed there.

Computational analyses, based on realistic models of pyramidal neurons with complex dendritic trees, suggest that for the most distal synapses on a pyramidal neuron, less than 10% of the synaptic charge will ultimately reach the soma. The same computational analyses, however, have shown that EPSPs in the soma are 100-fold smaller than the same EPSPs measured in distal dendrites, as observed in experiments (Fig. 11.2C and 11.2D). Why is there a 100-fold difference between dendritic and somatic EPSP amplitude if 10% of the synaptic charge reaches the soma? This is because a second factor influences the size of the EPSP at each location: the local input resistance (or impedance, the frequency-dependent equivalent of resistance). Small-diameter dendrites have very high input impedances, while the soma typically has a much lower input impedance, because of its larger size. Ohm's law dictates that a given amount of synaptic charge will produce a much larger EPSP in high-impedance dendrites than in the low-impedance soma, even if no charge were lost via alternative pathways.

These two effects—charge loss and difference in local input impedance—combine to produce very large differences between the amplitude of the EPSP at the synapse

BOX 11.2

DENDRITIC RECORDING

Ever since Santiago Ramón y Cajal looked down a microscope over a hundred years ago and observed that individual nerve cells have complex processes called dendrites neuroscientists have wondered at their properties. How dendrites influence the way neurons process synaptic input remained a mystery until the 1950s and 1960s, when theoretical models were used to explore dendritic function (Box 11.1). Experimental observations of dendritic properties had to wait until the 1970s and 1980s, when the first recordings were made from neuronal dendrites using sharp microelectrodes (Fig. 11.4). Due to their small size, these recordings, while very influential, were rare. Furthermore, because these recordings were performed "blind" (without visual information to guide microelectrode placement), identification of the site of recording was difficult, and recordings were restricted to brain regions where dendrites predominated.

The application of differential interference contrast optics to visualize neurons in brain slices allowed the possibility to record from neurons under visual control (Takahashi, 1978). This method was significantly enhanced by the introduction of the patch-clamp technique (Edwards, Konnerth, Sakmann, & Takahashi, 1989). Advances in optics, as well as the use of both infrared light, which reduces light scattering, and video microscopy, to enhance contrast, allowed dendrites to be visualized in living brain tissue for the first time (Dodt & Zieglgansberger, 1990). By combining these methods, it was possible to routinely record from dendrites under visual control (Stuart, Dodt, & Sakmann, 1993). This opened the door for direct experimental recordings to be made from dendrites using the power of the patch-clamp technique (Hamill, Marty, Neher, Sakmann, & Sigworth, 1981), allowing their properties and function to be investigated in detail. Using outside out and cell-attached recording, it has been possible to map the density and properties of voltage- and ligand-gated channels in the dendrites of different neuronal types. Furthermore, since both the dendrites and the cell body of neurons can be visually identified simultaneously, whole-cell recordings can be made from multiple locations on the same neuron, allowing the propagation of a variety of voltage signals in dendrites to be directly examined (Fig. B11.2).

As described in this chapter, these technical developments have led to a number of important discoveries about dendritic function in a variety of neuronal cell types, including direct observations of dendritic membrane properties, the finding that action potentials actively propagate back into the dendrites of many but not all neurons, and the description of a variety of different types of dendritic spikes and how they influence neuronal output. Critical to an understanding of how these dendritic properties influence brain function will be to determine how dendrites work in the intact brain. While we are still some way off achieving this goal, dendritic whole-cell recordings have been obtained *in vivo* in anesthetized preparations (Waters & Helmchen, 2004), and somatic whole-cell recordings can be obtained in awake, behaving animals (Lee, Manns, Sakmann, & Brecht, 2006). Using these methods, combined with optical techniques (Box 11.3), offers the possibility to observe dendritic activity at high resolution using the patch-clamp technique in the working brain.

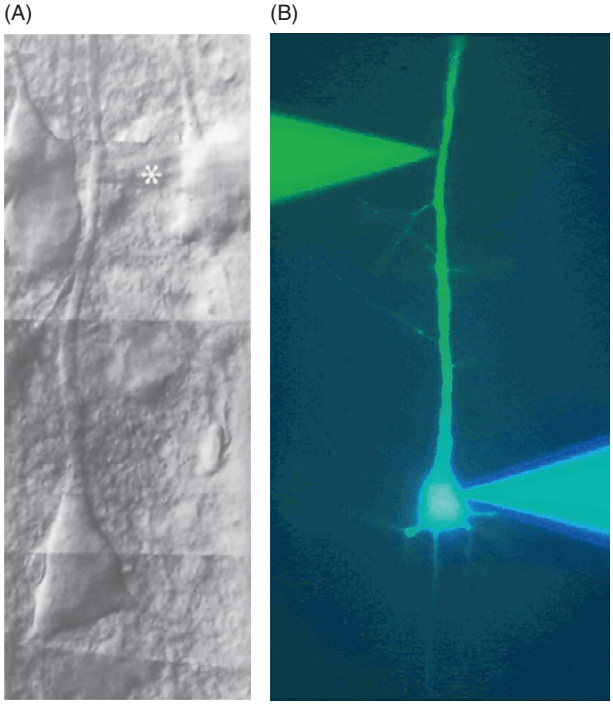


FIGURE B11.2 Examples of dendritic recording. (A) Differential interference contrast image of a cortical layer 5 pyramidal neuron in a living brain slice from rat somatosensory cortex. A patch pipette (asterisk) can be seen recording from the main apical dendrite. (B) Fluorescent image of a simultaneous somatic (blue) and dendritic (green) patch pipette recording from a cortical layer 5 pyramidal neuron. *Sources: (A) Adapted from Stuart, et al., 1993; (B) Adapted from Stuart and Sakmann, 1994.*

and at the soma. In passive dendrites, when comparing the ability of two synapses at different locations to depolarize the soma or axon (thereby influencing action potential initiation), only the first effect is relevant, because it is the amount of charge reaching the soma that matters. Distal synapses are located on small dendrites with high input impedance, which magnifies the local voltage change, thus resulting in very large local

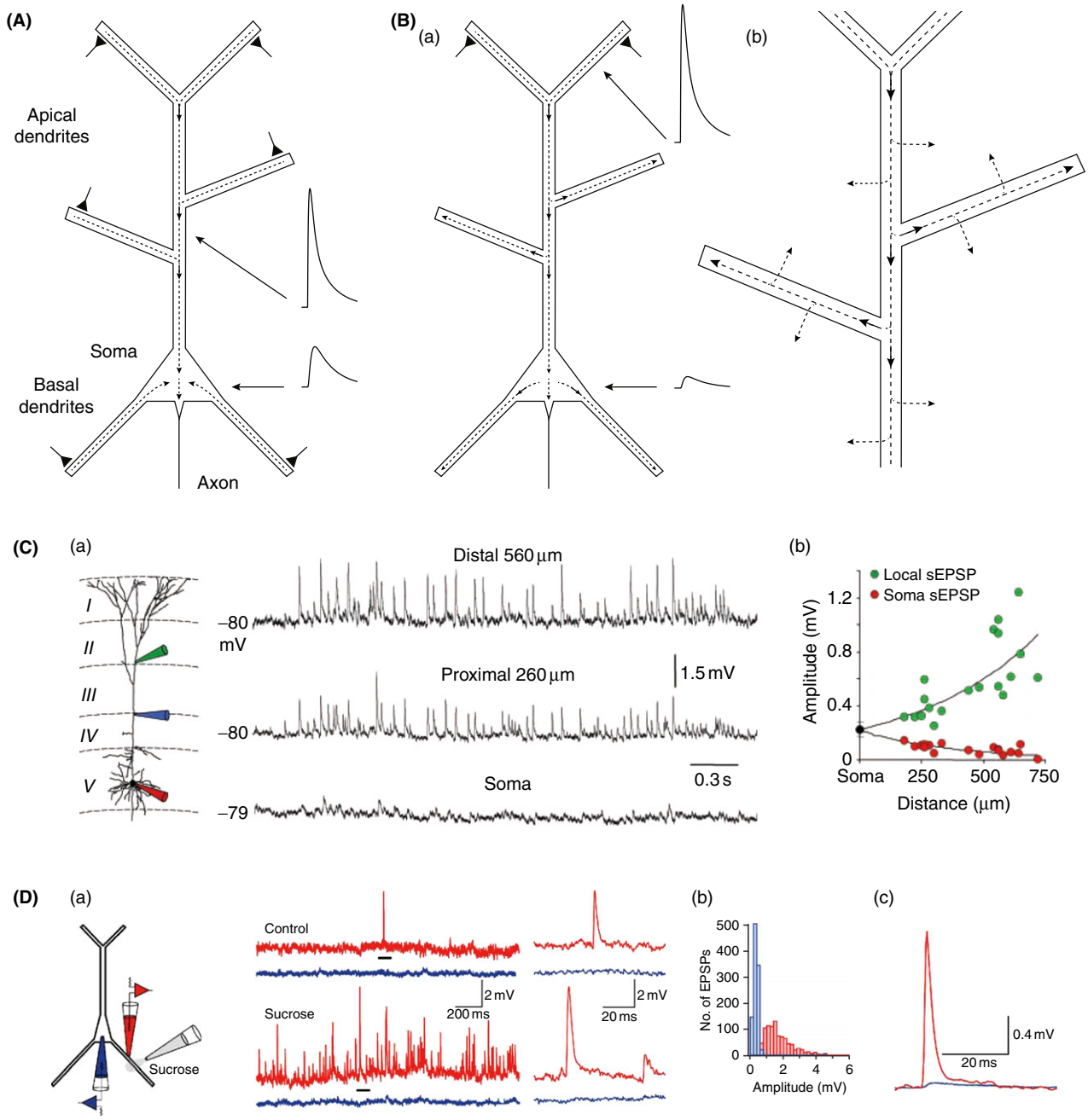


FIGURE 11.2 EPSPs are larger in dendrites than in the soma. (A) Schematic diagram of a pyramidal neuron illustrating synaptic current from the dendrites converging at the soma. Triangles indicate active excitatory synapses. Dashed lines indicate axial dendritic current, with arrowheads indicating direction of current flow. Schematic showing attenuation of EPSP from dendrite to the soma. Dendritic EPSP is larger than somatic EPSP because of higher input impedance of dendrite than soma and loss of charge between dendrite and soma. (B) (a) Similar to A, but showing different current flow when only distal apical synapses are activated. Dendritic EPSP is larger than in A because of higher input impedance of distal dendrite. Somatic EPSP is smaller than in A both because of fewer synapses and maximal loss of charge from distal dendrites. (b) Magnified view showing that some of the current flowing in dendrites is also lost across the dendritic membrane leak. Some membrane current would also flow in A. (C) (a) Triple whole-cell patch-clamp recordings from the soma, proximal dendrite, and distal dendrite. Spontaneous EPSPs were measured simultaneously at all three recording locations. (b) Amplitude of spontaneous EPSPs (sEPSP) measured locally and at the soma, as a function of sEPSP generation (i.e., the dendritic recording location at which the sEPSP was largest). (D) (a) Dual whole-cell patch-clamp recordings from the soma (blue) and a basal dendrite (red, 124 µm from soma) of a layer 5 neocortical pyramidal neuron. Sucrose was applied to evoke unitary EPSPs (by enhancing vesicle release), which are measured in the dendrite and soma. Traces at right are expansions of regions indicated by black bars. (b) Histogram of EPSP amplitudes measured in the soma (blue) and dendrite (red). (c) Average of EPSPs measured in the soma and dendrite. *Sources: (A, B) N. Spruston; (C) Williams and Stuart, 2002; (D) Nevian et al., 2007.*

EPSPs, but in passive dendrites this does not increase the efficacy of a synapse on action potential generation, because irrespective of whether a synapse is on the soma or on a distal dendrite, the amplitude of the somatic EPSP is determined solely by the synaptic charge that reaches the soma and the somatic input impedance. In the case of passive dendrites, only the extent to which synaptic charge is lost as current flows from the synapse to the soma affects the final impact of a distal synapse on action potential output. Thus, in the example above, the distal dendritic synapse would be only 10-fold less effective at depolarizing the soma than the same synapse located on a very proximal dendrite. As a result, it is important to realize that, in the passive case, the larger local voltage produced by more distal synapses does not in any way compensate for the distal location of these synapses, since this does not increase the amount of charge that flows into the cell at the synapse. In fact, the larger local voltage change in distal dendrites will reduce the electrical driving force for current flow at the synapse or, alternatively, may lead to activation of voltage-gated channels—important effects that are discussed later in the chapter.

Effects of Synaptic Kinetics on Dendritic Filtering

Another factor that affects the difference in amplitude of synaptic potentials in the dendrites versus the soma is the kinetics of synaptic current. As discussed in Chapter 6, current flow across the membrane is frequency dependent because of the membrane capacitance. Charge that is deposited onto the membrane capacitance can reach the soma eventually, but it is delayed. Thus, although the total amount of charge reaching the soma is not dependent on the kinetics of the synaptic current, the peak of the EPSP is attenuated more for fast synaptic potentials, such as EPSPs mediated by synchronous activation of AMPA-type glutamate receptors (GluRs), than for slow synaptic potentials, such as EPSPs mediated by activation of NMDA receptors.

The time course of synaptic potentials is faster in the dendrites than for a similar synaptic input to the soma (Fig. 11.3A); the time course of an EPSP in a dendritic spine is expected to be even faster. Near the synapse, membrane potential rises rapidly during synaptic current, owing to the rapid onset of synaptic currents. The decay of the dendritic EPSP is also quite fast as a result of current flowing axially along the dendrites, away from the synapse (Fig. 11.2B). The somatic EPSP is much slower, however. Its rise time is slower because of the delayed arrival of synaptic current from the dendrites and its decay, like the terminal decay of the dendritic EPSP, is relatively slow because it is governed by the membrane time constant (Fig. 11.3A), which is relatively slow because of the high membrane resistance and capacitance in most neurons (see Chapter 6). These factors combine to produce relatively slow EPSPs at the soma, resulting in a longer time window for summation of synaptic potentials in the soma than in the dendrites.

A special case that needs to be considered is experiments where an attempt is made to measure synaptic currents using voltage clamp. In order to obtain accurate measures of the amplitude and kinetics of synaptic currents, it is essential that the activated synapse be located very close to the voltage-clamp electrode. When attempting to voltage clamp synapses on distal dendrites using a somatic electrode, the resultant errors are very large (so-called "space-clamp" errors). Not only are the amplitude and kinetics of the synaptic current affected by the same factors as those described above for the filtering of synaptic potentials, but an additional error is caused by the difficulty of actually clamping the synaptic membrane potential using a somatic electrode. The distance between the somatic electrode and the synapse makes it impossible for the voltage clamp to control the voltage at the synapse, resulting in "voltage escape" at the synapse. This problem is greatest for fast synapses. Space-clamp errors—caused by loss of current from synapse to soma and escape of the synaptic potential from the voltage clamp—combine to produce very large distortions of the amplitude and kinetics of synaptic currents. The magnitude of these errors has been predicted and estimated using computational models but also validated using direct experimental measurements involving simultaneous dendritic and somatic recordings.

Interactions between Synapses in Passive Dendrites

If synapses were only activated one at a time, most cells would never fire an action potential. This is because the concerted action of many excitatory synapses is required to span the gap between the resting potential and action potential threshold. This gap is typically around 20 mV, so with an average EPSP amplitude of 0.2 mV, we can estimate that about 100 excitatory synapses would need to be activated to generate an action potential. This is at best a crude estimate, however, because it fails to take into account the influence of complexities such as inhibition, dendritic filtering, changes in driving force, and dendritic excitability. It is clear that interactions that occur between synapses in dendrites play a critical role in determining the impact of EPSPs on action potential output. While these interactions can be very complicated, a few simple principles can be described.

The first important point is that excitatory synapses summate in a sublinear fashion in passive dendrites

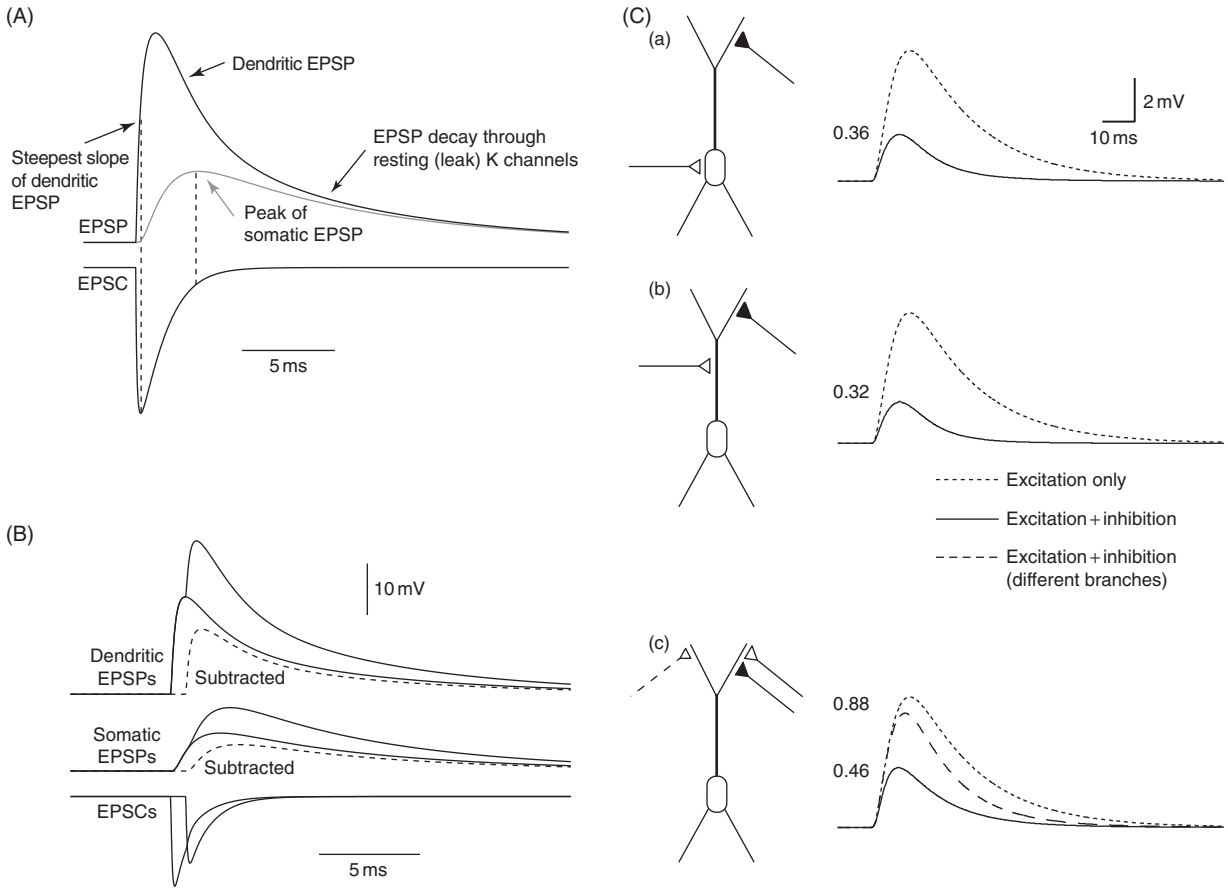


FIGURE 11.3 Synaptic interactions in passive dendrites. (A) Simulations showing the temporal relationship between the EPSP and the EPSC. Note the steepest slope of the dendritic EPSP occurs at the peak of the EPSC. The peak of the somatic EPSP occurs when the EPSC is very small (equal to leak current). The terminal decay of the EPSP (somatic and dendritic) is governed by the membrane time constant, but the initial decay of the dendritic EPSP is faster because of axial flow of current in the dendrites, away from the synapse. Amplitude scale is arbitrary. (B) Nonlinear summation of EPSPs in passive dendrites. Simulation of two excitatory synapses activated 0.75 ms apart. EPSPs and EPSCs are shown for activation of one synapse and activation of both synapses. Dashed line: arithmetic difference between EPSP for two synapses and one synapse. The single EPSP is about 20 mV; thus, the driving force for the second EPSC is reduced by about 33%, and the EPSPs and EPSCs are accordingly smaller. (C) Simulations of the effects of inhibition on a simple pyramidal cell model. Solid lines indicate response to excitation and inhibition. Inhibition slightly precedes excitation, but has a reversal potential equal to the resting potential, thus not producing any hyperpolarization on its own (i.e., only shunting inhibition). Dotted lines are excitation alone; dashed line is excitation and inhibition for alternate location of inhibition. Numbers indicate ratio of inhibited/uninhibited EPSP. (a) Distal excitation and somatic inhibition. (b) Distal excitation and on-path dendritic inhibition. (c) Distal excitation and distal dendritic inhibition, either on the same branch as excitation or on a different branch. *Sources: (A, B) Simulations by W.L. Kath and N. Spruston; (C) Spruston et al., 2008.*

(Fig. 11.3B). The reason for this is simple. If the resting potential of a neuron is –60 mV and the reversal potential of an excitatory synapse is 0 mV, then each synapse has a driving force of 60 mV. However, if enough synapses are activated to depolarize the dendrite to –40 mV, for example, the next activated synapse will have a driving force of 40 mV, which is 33% less than if that synapse were activated on its own. In other words, when multiple excitatory synapses are activated, each synapse has a reduced driving force compared to when it is activated on its own. It is not instructive to think of excitatory synapses as inhibiting each other because they always depolarize the neuron toward threshold for action potential firing; they can never do otherwise, but in passive dendrites the effect of activating multiple excitatory synapses is always less than the sum of the effects of each synapse individually.

A second important principle is that the magnitude of this sublinearity depends on the distance between the synapses. The same physical principles that govern the attenuation of EPSPs as they propagate to the soma also influence the extent of sublinear summation of activated synapses, as the depolarization experienced by one synapse depends on its distance from the other activated synapses. If the two synapses are close together, they have a strong influence on each other's driving force, and their summation will be

sublinear. On the other hand, if the synapses are far apart on the dendritic tree, each affects the other only minimally, so the net effect of activating both synapses will be closer to their linear sum.

When the interactions between synapses include inhibition, a third important principle is that "on-path" inhibition is more effective than "off-path" inhibition (Fig. 11.3C). Here, the "path" runs from the excitatory synapses in the dendrites to the action potential initiation zone in the axon. In order to effectively inhibit action potential initiation, inhibitory synapses must reduce the current that would have otherwise depolarized the action potential initiation zone. This happens best for on-path inhibition.

When considering inhibition, it is important to realize that inhibitory synapses can be effective even if they produce little or no hyperpolarization of the membrane potential on their own. This is often referred to as "shunting inhibition." This occurs most obviously when the reversal potential of an inhibitory synapse is the same as the resting potential. When activated on its own, such an inhibitory synapse has no effect on the membrane potential, because there is no driving force for current flow through the activated channels. When activated together with excitatory synapses, however, the inhibitory synapse shunts the current from the excitatory synapse, thus reducing the resultant somatic depolarization. In this scenario, the depolarization from the excitatory synapse creates a driving force for current flow through the inhibitory synapse. As a result, the depolarizing effect of Na^+ ions entering through excitatory channels (e.g., AMPA receptors) is reduced by the hyperpolarizing effect of Cl^- ions entering through the inhibitory channels ($GABA_A$ receptors).

Functional Implications of Dendritic Filtering

The fact that synaptic potentials are filtered between the synapse and the soma, affecting both their amplitude and time course, has important functional implications for synaptic integration. As mentioned above, excitatory synapses located on distal dendrites of large neurons (such as pyramidal neurons) are expected to produce somatic EPSPs that are less than one-tenth that of their more proximal counterparts. This means that in many neurons, such as pyramidal neurons of the CNS, activation of a single distal synapse could have an almost negligible effect at the soma.

Although distal synapses are expected, based on theoretical considerations, to have reduced impact at the soma compared to more proximal synapses, it is possible that at least some distal synapses might compensate for this disadvantage by having more AMPA receptors than more proximal synapses. While there is some experimental evidence that this is the case in CA1 pyramidal neurons, this is not a general principle of synaptic organization. In fact, reconstruction of synapse morphologies using serial-section electron microscopy has indicated that most (~80%) of the synapses on CA1 pyramidal neurons are very small, and immunostaining has indicated that these small synapses have very few AMPA receptors. The other 20% of the synapses have much larger PSDs and a higher density of AMPA receptors. These synapses are likely to have a larger impact at the soma. Evidence exists that they are more numerous in distal dendrites than in proximal dendrites, but they are nevertheless a minority. What is the purpose of covering dendrites with a large number of synapses that are very small? While it is true that if enough (e.g., hundreds) of these weak synapses are activated, they could indeed result in action potential initiation in the axon, it seems equally likely that these distal synapses have a different role. One possibility is that these synapses represent a large pool of synapses that can be recruited for conversion to the more powerful class of synapses through synaptic plasticity such as LTP. Another possibility, however, is that even small synapses can have significant local effects, because of the magnification of local potential by the high impedance of small dendritic branches, thus resulting in activation of dendritic voltage-gated channels. For this reason, to answer the question of how distal synapses influence the processes of synaptic integration and action potential generation, it is necessary to consider the more complex case of synaptic integration in active dendrites.

SYNAPTIC INTEGRATION IN ACTIVE DENDRITES

The foregoing discussion was based on considering how synaptic potentials are integrated in passive dendrites. While this serves as a good starting point for understanding synaptic integration, another layer of complexity exists: in every dendrite where neuroscientists have looked, voltage-gated channels have been found. Among the most important of these dendritic channels are voltage-gated Na^+ channels (Na_v), Ca^{2+} channels (Ca_v), and K^+ channels (K_v). These channels have been observed in dendrites using immunostaining of the channels with both light and electron microscopy, and their functional properties have been studied using dendritic patch-clamp recording and imaging. The identification of these channels in dendrites using modern methods that allowed direct observation of the channels validated earlier studies that inferred the excitability of dendrites from the properties of extracellular field potentials.

Intracellular recordings have also been used to study dendritic excitability. Initially, these recordings were performed using sharp microelectrodes (Fig. 11.4). More recently, however, whole-cell patch-clamp recordings

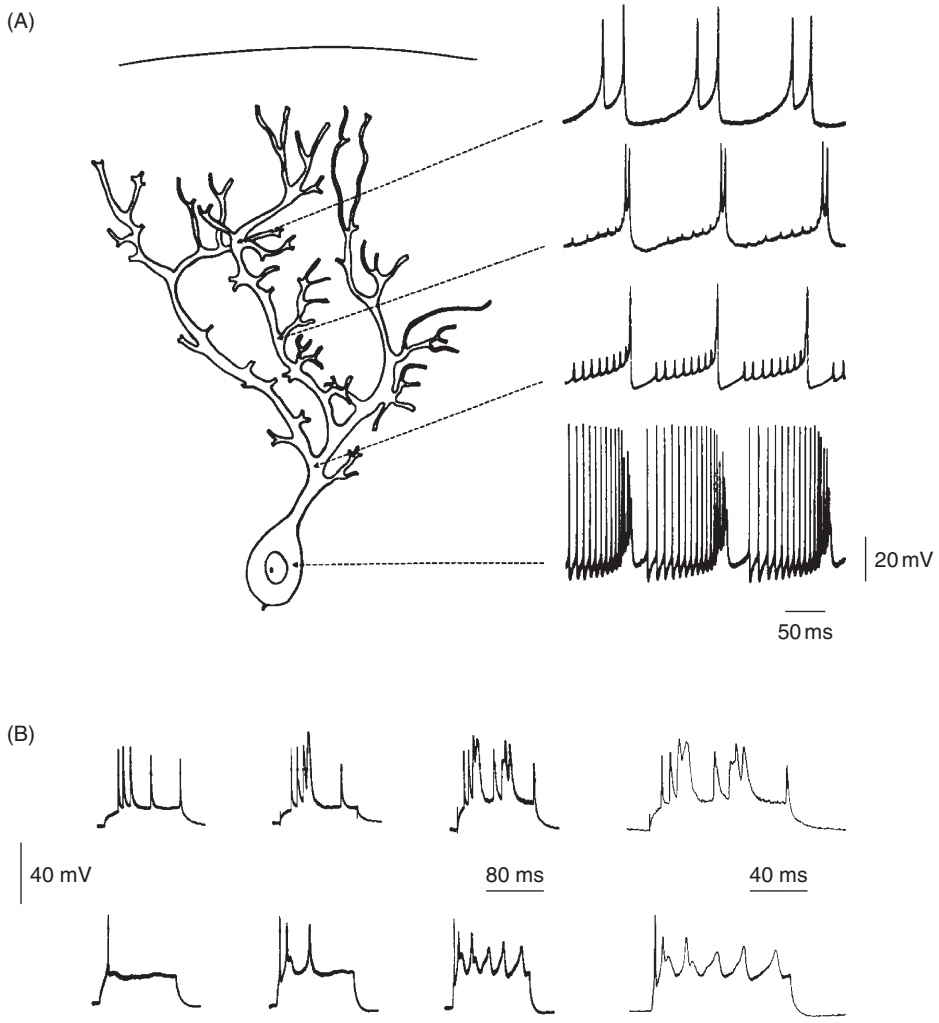


FIGURE 11.4 Early examples of dendritic recordings. (A) Dendritic and somatic recordings obtained from cerebellar Purkinje neurons (composite of separate recordings). Large action potentials at the soma are severely attenuated in the dendrites. Distal dendritic recordings show large, slow spikes mediated by Ca_v channels. (B) Recordings from dendrites of CA1 pyramidal neurons in intact slices (top) and after separation of dendrites from somata (bottom). All traces are responses to current steps. The rightmost trace is a temporally expanded view of the response to the left. *Sources: (A) Llinás and Sugimori, 1980; (B) Benardo et al., 1982.*

have been obtained from dendrites (Box 11.2). Most of these experiments have been performed using *in vitro* brain-slice preparations, but some dendritic recordings have also been obtained *in vivo*. In addition to single whole-cell recordings, simultaneous recordings have been performed from multiple locations on the same neuron, facilitating careful analysis of the relative size and timing of various events (e.g., EPSPs, IPSPs, spikes) in different parts of the neuron.

Dendritic excitability has now been studied in several cell types, but the ones most intensively studied are pyramidal neurons (hippocampal CA1 and neocortical layer 5 and layer 2/3) and cerebellar Purkinje cells. Studies of these and other neurons have revealed that the excitable properties of dendrites depend on the cell type. For example, CA1 pyramidal neurons are quite different from layer 5 pyramidal neurons, and Purkinje cells are even more different (as one might expect from the very distinct morphology of their dendritic tree). Despite differences between neuronal types, it is possible to extract a number of important general principles from these studies.

Weakly Excitable Dendrites and Backpropagating Action Potentials

One of the central principles is that dendrites are weakly excitable; that is, they contain voltage-gated channels that can support action potential propagation and dendritic spike generation, but they are not as excitable as the axon. The most direct evidence for this comes from studies involving simultaneous recording from the soma and a dendrite. These experiments have revealed that action potentials occur first in the axon and later in the soma and dendrites (Fig. 11.5). This is true even when the stimulus leading to the action potential is dendritic (e.g., activation of synapses near the dendritic recording electrode). This result has been interpreted to indicate that the action potential is initiated in the axon (in the axon initial segment), then "backpropagates" into the soma and dendrites. Indeed, experiments involving simultaneous recording from the axon and the soma have revealed that the axonal action potential precedes the somatic action potential, again confirming

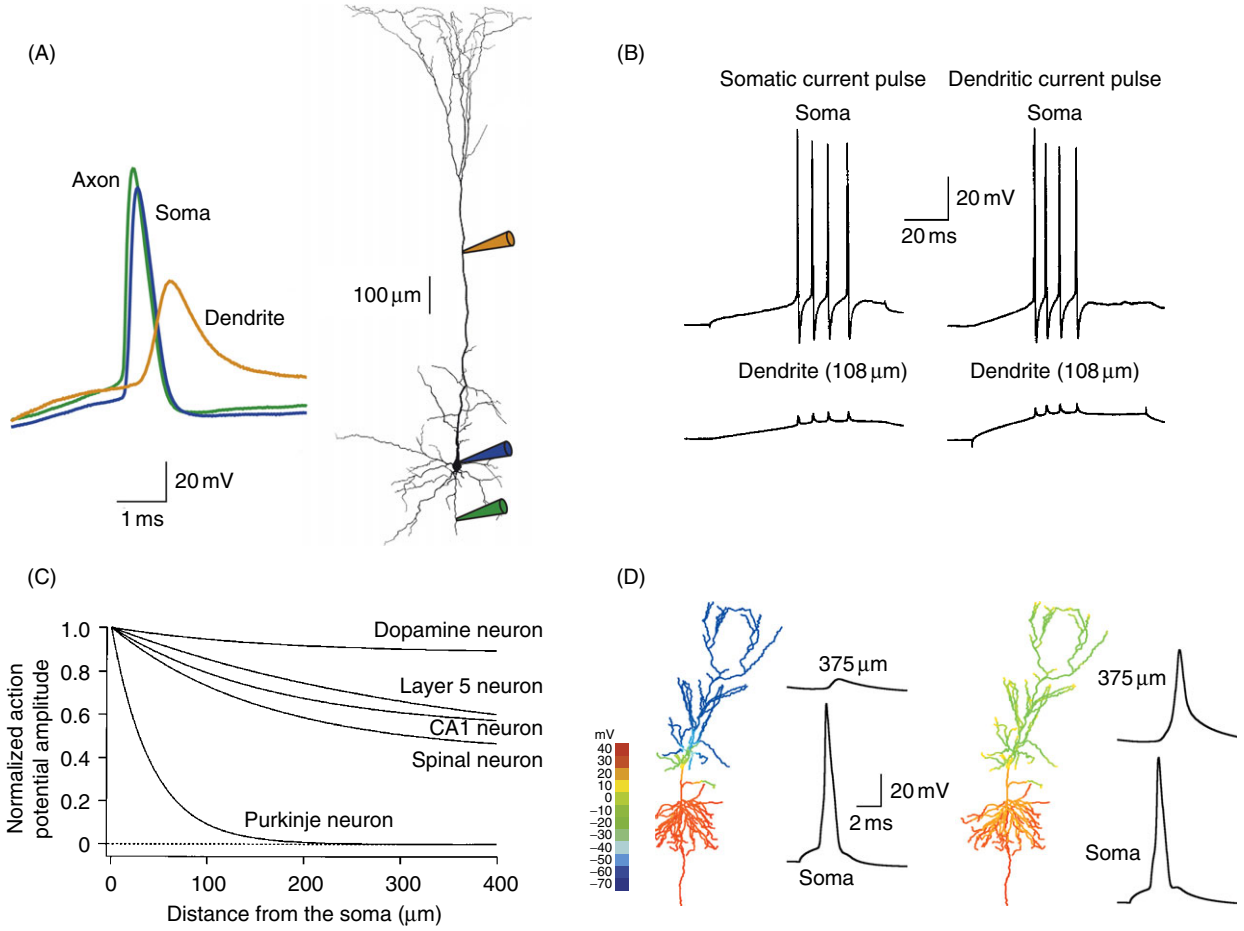


FIGURE 11.5 Backpropagating action potentials. (A) Triple recording from a neocortical layer 5 pyramidal neuron shows action potential initiation in the axon and backpropagation into the soma and dendrites. (B) Simultaneous dendritic recording from soma and dendrites of cerebellar Purkinje cell shows severe attenuation of action potentials recorded in the dendrites, regardless of the site of current injection. (C) Variable amplitude attenuation of backpropagating action potentials in different types of neurons. (D) Simulations of backpropagating action potentials in CA1 pyramidal neurons reveal sensitivity to Na_v and K_v channel density in the distal dendrites, which determines whether or not the action potential successfully invades the apical dendritic tuft. The model shown on the right has a higher density of dendritic Na_v channels than the model on the left. *Sources: (A) Häusser et al., 2000; (B) Stuart and Häusser, 1994; (C) Stuart et al., 1997b; (D) Golding et al., 2001.*

earlier conclusions based on extracellular recordings. Thus, the axon appears to have a much lower threshold for action potential initiation than the dendrites. The low axonal threshold results from the fact that, compared to the dendrites, the axon has a much higher density of Na_v channels and possibly a higher density of K_v channels.

Another observation indicating that dendrites are only weakly excitable is that backpropagating action potentials tend to attenuate as they propagate (Fig. 11.5). Dendritic recordings have revealed that the amplitude of the backpropagating action potential decreases with distance from the soma. The steepness of this attenuation (amplitude as a function of distance), however, varies between cell types. Pyramidal neurons have intermediate attenuation, while Purkinje cells exhibit much greater attenuation and dopamine neurons in substantia nigra have almost no attenuation. The degree of attenuation depends on two factors. One is the density and ratio of Na_v and K_v channels, and the other is the amount of branching in the dendritic tree. Branch points create impedance mismatches and capacitance loads, which means more current is required to support reliable propagation of the action potential through each branch point. This effect can be small for individual branch points, but the cumulative effect of many branch points can be very significant. For example, in cerebellar Purkinje cells, there are so many branch points that active action potential backpropagation is completely inhibited. Though this is in part due to the low density of Na_v channels in Purkinje cell dendrites, simulations have shown that very high densities of Na_v channels would be required to overcome the extensive branching in these dendrites.

Most of our knowledge about the amplitude of backpropagating action potentials at different distances from the soma comes from dendritic patch-clamp recording. Because these recordings have mostly been made from relatively large-diameter dendrites, less is known about the amplitude of backpropagating action potentials in smaller dendrites, such as basal, apical oblique, or apical tuft branches in pyramidal neurons or the fine branchlets of Purkinje cells. It is extremely difficult to place electrodes on such small dendrites, so imaging is a valuable complement to electrophysiological recording (Box 11.3). Indeed, imaging experiments have provided experimental support for the notion that even small-diameter dendrites are excitable.

Nonuniform Distribution of Voltage-Gated Channels in Dendrites

Another important principle of dendritic excitability is that the expression of voltage-gated channels is not uniform across the dendritic tree. The two best-characterized examples of this are the A-type K_v channels (K_v4) in CA1 pyramidal neurons and the hyperpolarization-activated cation (HCN) channels in CA1 and layer 5 pyramidal neurons. Patch-clamp recordings from these currents have shown that both channel types have much higher densities in distal apical dendrites than in the soma or proximal apical dendrites. Immunostaining has verified and extended these findings, but important questions

BOX 11.3

IMAGING DENDRITIC FUNCTION

The microscope was for a long time the only method by which experimentalists could gain information about the properties of dendrites, though the information yielded by light and electron microscopes was initially purely structural. The advent of dendritic microelectrode recording in the 1970s, and subsequently patch-clamp recording from dendrites in the 1990s (see Box 11.2), transformed our view of the functional properties of dendrites, offering us direct information about the electrical events taking place within the dendritic tree. Over the last 15 years, however, a variety of new imaging techniques have emerged as alternatives to electrophysiological approaches for investigating the functional properties of dendrites. Using light to probe function has several key advantages over conventional electrophysiological techniques (Scanziani & Häusser, 2009): light is noninvasive in that it normally does not interact with neurons; it can penetrate deep into tissue; offers outstanding spatial resolution; can permit detection of single molecular events; allows simultaneous measurement from multiple locations; and, crucially (with the advent of genetic and subcellularly targetable probes), permits recording from specific cell types and precisely defined subcellular compartments.

Early approaches for harnessing the power of imaging to probe the functional properties of dendrites involved the use of fluorescence microscopy, using confocal microscopes, CCD cameras, or photodiode arrays. In conjunction with the use of sensitive calcium indicators, introduced into the cell by microelectrodes, these approaches permitted the investigation of calcium signaling in distal dendrites, and even individual dendritic spines, to be probed for the first time. Calcium has often been used as a surrogate for electrical activity, particularly in compartments inaccessible to recording electrodes, but with the caveat that Ca^{2+} signals are affected not just by membrane potential, but also by other factors, including channel density, surface-to-volume ratio, and Ca^{2+} buffering and transport. Of course, calcium dynamics have also been studied in their own right due to the immense importance of calcium as a biochemical signalling molecule. The advent of improved voltage-sensitive dyes, together with faster and more sensitive CCD cameras, has more recently enabled voltage-sensitive dye measurements to be made directly from dendrites, allowing direct measurements of voltage in small dendritic compartments (though with limitations due to intracellular compartmentalization of dye; lack of calibration in distal dendrites; and problems with dye toxicity).

A fundamental breakthrough in using light to probe dendritic function came with the invention of two-photon microscopy (Denk & Svoboda, 1997). This approach harnesses the highly nonlinear nature of two-photon absorption to limit excitation to a tiny (femtoliter) volume, permitting high-resolution imaging deep within tissue (Fig. B11.3). Two-photon microscopy has been applied to study the structural and functional properties of dendrites and spines both in brain slices and more recently also *in vivo* (see Figure; Denk & Svoboda, 1997). It has led to numerous important discoveries both about the morphological and functional dynamics of dendritic spines as well as information processing within the dendritic tree. Furthermore, two-photon microscopy allows patch-clamp recordings to be targeted to particular cell types, and even to individual dendrites, and also allows recording locations to be verified with high precision, both in brain slices and *in vivo*, and has thus become an indispensible tool for dendritic physiologists. Recent developments in the miniaturization of two-photon optics have enabled imaging of neuronal calcium signals in freely moving animals equipped with head-mounted two-photon microscopes.

BOX 11.3 (*cont'd*)

Another important optical approach for probing dendritic function is the use of two-photon excitation for uncaging of neurotransmitters, which in conjunction with appropriate cages now allows individual glutamatergic and GABAergic synapses on dendrites to be activated with high temporal and spatial precision. This approach has allowed the input-output function of individual dendrites to be probed with unrivalled temporal and spatial precision (e.g., Branco, Clark, & Häusser, 2010) and has also permitted the investigation of dendritic plasticity rules.

There is currently tremendous enthusiasm about the use of imaging techniques for examining dendritic function, driven by advances in microscope resolution (Hell, 2007), innovative sample preparation (Micheva & Smith, 2007), and development of new sensors (Miyawaki, 2011). In particular, the advent of new genetically encoded probes for calcium, voltage, and a variety of signal transduction molecules, as well as the ability to target these probes to specific cell types and/or subcellular compartments, promise to offer us unprecedented insights into both electrical and chemical signaling in the most distal corners of dendritic trees. While electrophysiological approaches will probably always maintain their status as the "gold standard" for investigating electrical signaling in dendrites, due to their superior sensitivity and temporal resolution (Scanziani & Häusser, 2009), it is clear that optical solutions for probing dendritic function will become increasingly popular.

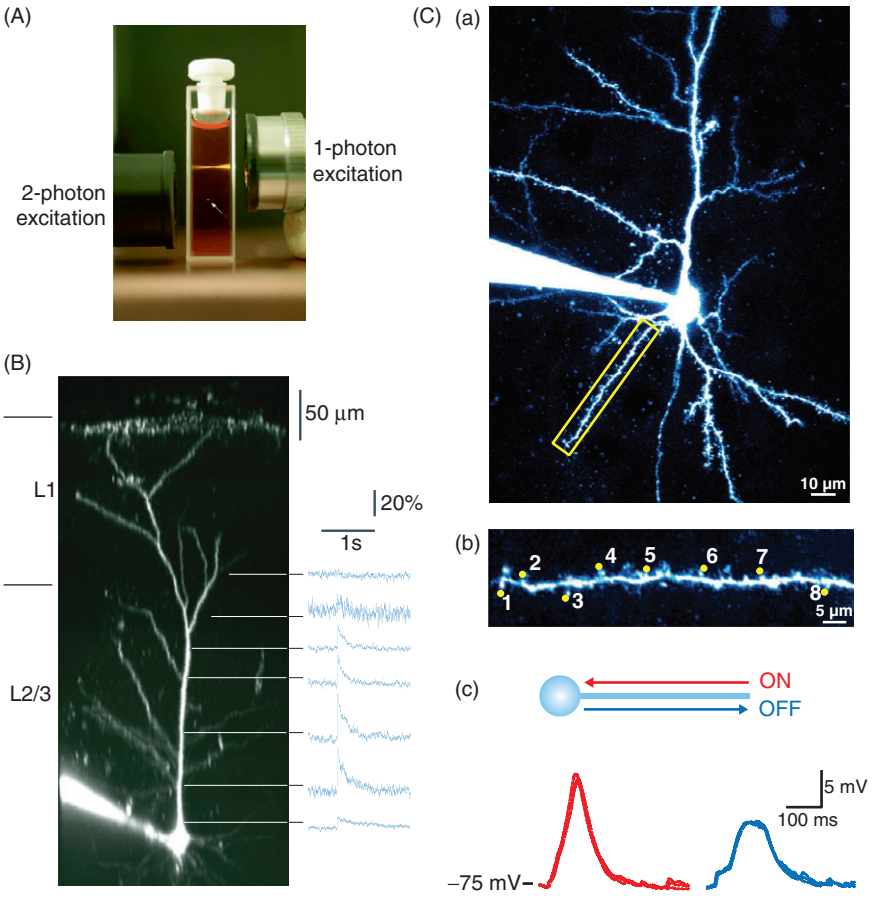


FIGURE B11.3 Using optical techniques to probe dendritic function. (A) A cuvette filled with a fluorescent dye is illuminated with either one-photon or two-photon excitation (top and bottom objectives, respectively). Note the remarkable spatial precision of the two-photon excitation. (B) *In vivo* imaging of dendritic calcium signaling. Left, a single layer 2/3 pyramidal cell in an anaesthetized rat is filled with a calcium dye (Oregon Green BAPTA-1) via a somatic patch-clamp recording and imaged using 2-photon microscopy. Right, Ca^{2+} transients (averages of 3–5 trials) that were evoked by single backpropagating action potentials at the positions indicated along the dendrite. (C) Probing computational properties of single dendrites using two-photon glutamate uncaging. (a) A pyramidal cell in a neocortical slice is filled with a fluorescent dye via a somatic patch pipette. (b) Spines along a single dendrite are imaged using two-photon microscopy, and the dendrite is bathed with a caged glutamate compound (MNI-glutamate). Synaptic transmission is then mimicked by brief two-photon uncaging at individual spines (yellow dots). (c) The synapses are activated in a sequence, either moving towards the soma (ON, red trace), or away from the soma (OFF, blue trace), producing a directionally selective EPSP at the soma. *Sources: (A) Courtesy of Brad Amos, MRC, LMB, Cambridge; (B) Waters et al., 2003; (C) Branco et al., 2010.*

remain about the distribution of these and other channels. Numerous challenges must be overcome before we can determine the distribution and functional properties of voltage-gated channels in dendrites from which direct recordings are very difficult. These challenges include questions about antibody specificity, difficulty distinguishing proximal versus distal dendrites when they are overlapping (as in apical oblique branches of pyramidal neurons), or expression of channels in closely apposed presynaptic and postsynaptic structures (e.g., boutons and spines). Furthermore, expression does not always equate to function, so while staining methods can in principle tell whether or not a channel type is present in a particular dendrite, staining cannot be used to determine if the observed channels are functional and, if so, what their specific properties are.

The nonuniform expression of voltage-gated channels in dendrites is likely to have significant functional implications. This is best understood for the case of K_v4 channels in CA1 pyramidal neurons, where the increasing density at more distal locations contributes to the distance-dependent attenuation of backpropagating action potentials and raises the threshold for generation of dendritic spikes in these neurons (see below). Because there are so many different types of channels (e.g., low- and high-threshold Ca_v channels; Ca^{2+}-activated K^+ channels), determining the subtypes, distribution, and properties of these channels in dendrites of different types of neurons will be critical to a better understanding of how synaptic inputs are integrated in the brain.

Dendritic Spikes

Although the lowest threshold for action potential initiation is in the axon, most dendrites can generate dendritic spikes (Figs. 11.6 and 11.7). Because of the weakly excitable nature of the dendritic tree, dendritic spikes have a relatively high voltage threshold. Recall, however, that dendrites locally magnify synaptic potentials because of their high input impedance. Thus, even a small number of synapses can produce local depolarizations that are large enough to activate voltage-gated channels in the dendrites.

One consequence of the weakly excitable nature of dendritic trees is that dendritic spikes do not propagate reliably throughout them. Like the backpropagating action potential, the attenuation of forward propagating dendritic spikes depends both on the types of channels activated during the spike and on the branching of the dendritic tree. Because of the structure of branch points, however, forward propagation is even less reliable than backpropagation. The difference arises from the geometry of the dendritic tree. Consider a backpropagating action potential traveling from the main apical trunk into an oblique branch in a pyramidal neuron. The amount of current required to generate the action potential in the trunk is quite large because of the large size of this dendrite. The oblique branch is much smaller, so a small fraction of the charge in the trunk is able to depolarize the oblique branch to threshold, thus resulting in successful backpropagation through the branch point. If an action potential is generated in the oblique branch and propagates toward the apical trunk, however, the situation is very different. The amount of charge required to generate a dendritic spike in the oblique branch is small because of the small diameter of the dendrite. As this charge flows into the apical trunk, there is too little current to depolarize the trunk to threshold, so the dendritic spike fails at the branch point. The asymmetry of propagation through dendritic branches is caused by an impedance mismatch: the apical trunk has a low input impedance and the oblique branch has a high input impedance. It is relatively easy for a regenerative event such as an action potential or dendritic spike to propagate from a low-impedance to a high-impedance structure but relatively hard for propagation to occur from a high-impedance to a low-impedance structure.

Dendritically initiated spikes come in different varieties: Na^+ spikes, Ca^{2+} spikes, and NMDA spikes (Figs. 11.6 and 11.7). These spikes are named for the dominant channel type activated during the spike, although in fact multiple types of channels are usually activated during each type of event. For example, even when a Na^+ spike is observed, NMDA receptor activation might be critical for providing the depolarization that triggers it, and the resulting depolarization from the spike is usually large enough to activate Ca_v channels. The spike is still called a Na^+ spike—typically identified as the fastest dendritic event (because of the fast kinetics of Na_v channels)—but a significant increase in intracellular Ca^{2+} would be observed in a Ca^{2+} imaging experiment because of the activation of NMDA receptors and Ca_v channels.

The first dendritically initiated spikes to be described were dendritic Ca^{2+} spikes, which were observed in experiments performed in the 1970s and 1980s using sharp microelectrodes to record from the dendrites of Purkinje neurons and pyramidal neurons (Fig. 11.4). These spikes were initiated in response to strong excitatory synaptic inputs and were presumed to have a dendritic origin because of their large amplitude in dendritic recordings. Subsequent to these early experiments, the dendritic origin of these Ca^{2+} spikes was confirmed in simultaneous somatic and dendritic recordings using patch-clamp recording (Box 11.2). Ca^{2+} spikes have a long duration (relative to Na^+ spikes), consistent with the slow kinetics of Ca_v channels compared to Na_v channels. In some cases, dendritic Ca^{2+} spikes of especially long duration have been called "plateau potentials." Although dendritic Ca^{2+} spikes were first observed only with strong synaptic stimulation, they have also been

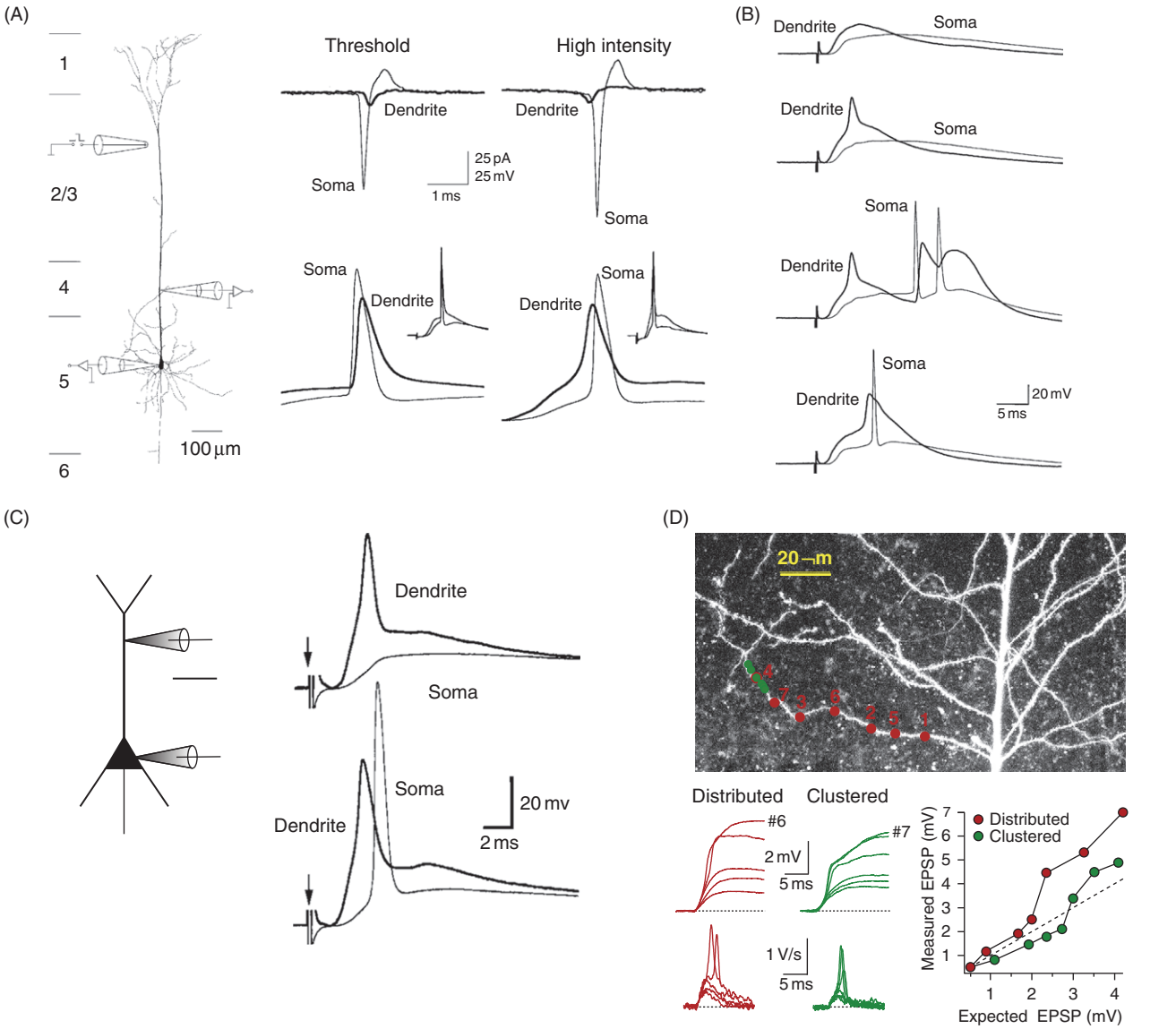


FIGURE 11.6 Dendritic sodium spikes. (A) Simultaneous somatic and dendritic (175 µm from soma) recordings from a neocortical layer 5 pyramidal neuron. Using synaptic stimulation at threshold for action potential initiation (left), the somatic action potential precedes the dendritic action potential, both in cell-attached (top) and whole-cell (bottom) recordings. At higher synaptic stimulus intensity (right) the action potential occurs first in the dendrites. (B) Similar recording configuration with dendritic recording 440 µm from the soma. Example traces show (top to bottom): subthreshold EPSP, isolated dendritic spike, dendritic spike followed by somatic burst, dendritic spike preceding somatic action potential. (C) Somatic and dendritic recordings in a hippocampal CA1 pyramidal neuron showing an isolated dendritic spike (top) and a dendritic spike preceding a somatic action potential (bottom). (D) Two-photon uncaging of glutamate onto the dendrites of a CA1 pyramidal neuron. Somatic recordings show a nonlinear increase in the somatic EPSP amplitude and slope, indicative of a dendritic spike that did not trigger an action potential in the axon or soma. *Sources: (A, B) Stuart et al., 1997a; (C) Golding and Spruston, 1998; (D) Losonczy and Magee, 2006.*

shown to occur in the apical tuft of pyramidal neurons when weaker synaptic excitation is paired with one or more backpropagating action potentials. These events have been called "backpropagating action potential evoked Ca^{2+} spikes" or "BAC spikes" (Fig. 11.7). They are of particular interest because they may signal the coincidence of synaptic input to the apical tuft and other dendritic domains of pyramidal neurons.

Dendritic Na^+ spikes are fast events (compared to Ca^{2+} spikes), so they tend to be very localized and do not propagate reliably out of the branch in which they were generated (Fig. 11.6). This fact is likely responsible for the relatively late discovery of dendritic Na^+ spikes, compared to Ca^{2+} spikes. Dendritic Na^+ spikes were first definitively identified when it became possible to simultaneously monitor somatic and dendritic membrane potential using

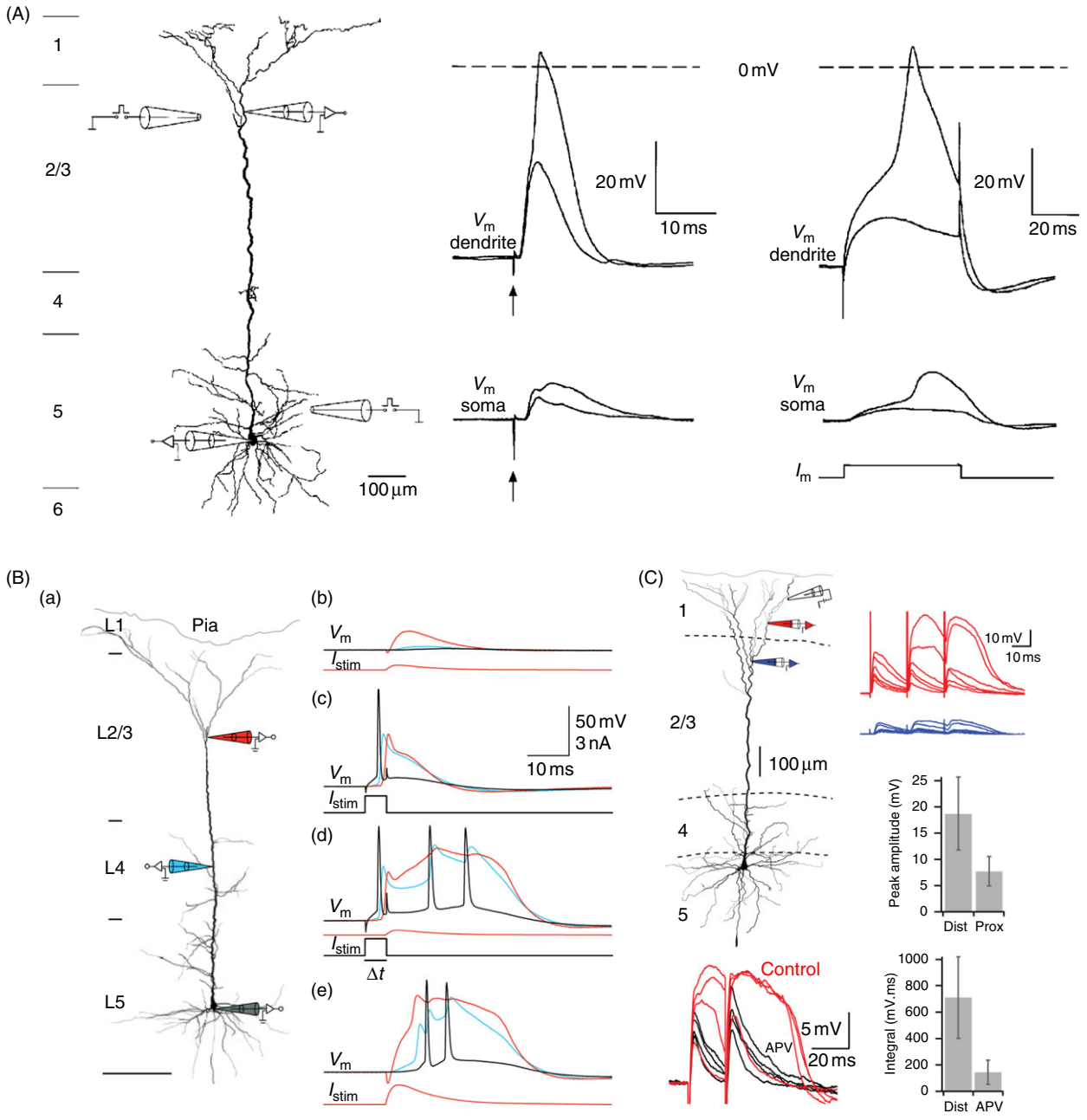


FIGURE 11.7 Dendritic calcium spikes and NMDA spikes. (A) Simultaneous somatic and dendritic (920 µm from soma) recordings from a neocortical layer 5 pyramidal neuron. Synaptic stimulation in layer 2/3 (left traces) or current injection into the dendrite (right) resulted in a dendritic Ca^{2+} spike (subthreshold responses also shown) in the dendrites that did not propagate to the soma (soma, 440 µm and 770 µm from soma). (B) (a) Schematic of triple recording from a layer 5 pyramidal neuron. (b) Responses to EPSC-like dendritic current injection. (c) Responses to somatic current injection. (d) Responses to combined somatic and dendritic current injections. (e) Responses to larger dendritic current injection. (C) Dual recording 715 µm and 875 µm from the soma in a layer 5 pyramidal neuron. Synaptic stimulation in layer 1 yields dendritic spikes that are restricted to the distal dendritic recording location and blocked by the NMDA receptor antagonist APV. *Sources: (A) Schiller et al., 1997; (B) Larkum et al., 1999; (C) Larkum et al., 2009.*

patch-clamp recording in brain slices. This technique allowed definitive identification of dendritic spikes that occurred either before or in the absence of a somatic action potential. At first, dendritic Na^+ spikes were only observed in response to relatively strong stimulation of excitatory synapses. The strong stimulus used to evoke dendritic spikes suggested that they might only occur under conditions of extreme excitation, thus raising questions about their physiological significance. However, subsequent experiments using two-photon Ca^{2+} imaging

and glutamate uncaging (Box 11.3), combined with somatic and dendritic recording, have indicated that dendritic spikes may occur in response to clustered activation of a small number of synapses on a single branch (i.e., less than ten synapses). Dendritic Na^+ spikes are hard to detect, even with dendritic recording, because the spikes do not propagate reliably into the large-diameter dendrites from which dendritic recordings are typically obtained.

Dendritic NMDA spikes result from activation of NMDA receptor channels, which are voltage sensitive because of their block by Mg^{2+}. When glutamate is bound to NMDA receptors, they can produce a regenerative event, as depolarization leads to relief of Mg^{2+} block, resulting in more inward current and more depolarization. If enough NMDA receptors are activated, a regenerative spike can occur, even if significant numbers of Na_v and Ca_v channels are not activated. NMDA spikes have one important feature that is fundamentally different from Na^+ or Ca^{2+} spikes. NMDA spikes cannot propagate into regions of the dendrite where NMDA receptors are not activated by glutamate (or where they are activated only in small numbers). As noted above, dendritic Na^+ spikes also propagate poorly and typically remain restricted to the dendrite in which they were originated. Unlike NMDA spikes, however, dendritic Na^+ spikes can propagate into regions of the branch where synapses have not been activated. NMDA spikes cannot propagate in this way because of the additional requirement that the receptor needs to be activated by glutamate before it can respond to depolarization.

Functions of Dendritic Spikes

An important question regarding excitable dendrites is to determine the functions of dendritic spikes and backpropagating action potentials. Two key issues are which of these events occurs *in vivo*, and under what conditions? Although good progress is being made to develop the methods necessary to address these questions, they remain unanswered. Thus, most of what we can say about the function of dendritic spikes stems from what we have learned about them from *in vitro* brain-slice experiments. Generally, the possible functions of backpropagating action potentials and dendritic spikes can be divided into three categories: a role in synaptic integration, a role in local communication, and a role in plasticity.

Dendritically initiated spikes are likely to play a central role in the process of synaptic integration, but understanding how this works is not straightforward. For example, if a dendritic Na^+ spike does not propagate out of the dendritic branch in which it was generated, how can it influence action potential initiation in the axon? The most likely answer to this puzzling question is that the dendritic spike, although not propagating actively, does deliver additional charge to the axon, over and above the charge entering at synapses alone. In this way, dendritic excitability lowers the number of excitatory synapses required to initiate action potential firing in the axon. Thus, even though individual synapses in the distal dendrites minimally depolarize the axon (because of the dendritic filtering discussed above), the influence of distal synapses can be enhanced when they contribute to the generation of dendritically initiated spikes.

A leading theory in the field of dendritic integration is that dendritic spikes allow neurons to respond to coincident activation of multiple dendritic domains. For example, if Na^+ spikes or NMDA spikes are generated in two oblique branches of pyramidal neurons, enough charge can be delivered to the axon to generate an action potential, even though the dendritic spikes do not actively propagate all the way to the axon. Another example is the BAC spike, discussed above; when coincident input occurs to the apical tuft and the basal dendrites of pyramidal neurons, a BAC spike can be generated in the tuft, leading to a burst of action potentials at the soma. Yet another example from pyramidal neurons is that spikes generated in the apical tuft normally fail at the branch point between the tuft and the main apical dendrite. When this critical region near the branch point is depolarized by other synaptic inputs, however, the tuft spike can propagate actively along the main apical trunk and trigger action potential firing in the axon. Thus, dendritic spikes may allow the neuron to detect the coincident activation of excitatory synaptic inputs to different dendritic compartments. A central question, then, which is the focus of much ongoing work, is what kinds of inputs, carrying which kinds of information, arrive in the different dendritic compartments of pyramidal neurons.

The concept of coincidence detection through dendritic excitability is not restricted to pyramidal neurons. Cerebellar Purkinje neurons can also detect the coincidence of inputs via the parallel fibers and the climbing fiber. Although climbing fibers on their own are sufficiently strong to reliably evoke dendritic Ca^{2+} spikes in Purkinje cells, coincident input via the parallel fibers can increase the number of Ca^{2+} spikes evoked by a single presynaptic action potential in the climbing fiber. Furthermore, the Purkinje cell dendritic tree branches so extensively that Ca^{2+} spikes can cause larger depolarizations in the vicinity of the activated parallel fibers than in other portions of the dendrites. One important outcome of the larger dendritic spikes associated with such coincident activity is the induction of long-term depression (LTD) of the activated parallel fibers.

The end result of coincidence detection is usually a dendritic spike. As an outcome of synaptic integration, a dendritic spike can enhance action potential firing in the axon. In the case of Ca^{2+} spikes (including BAC spikes),

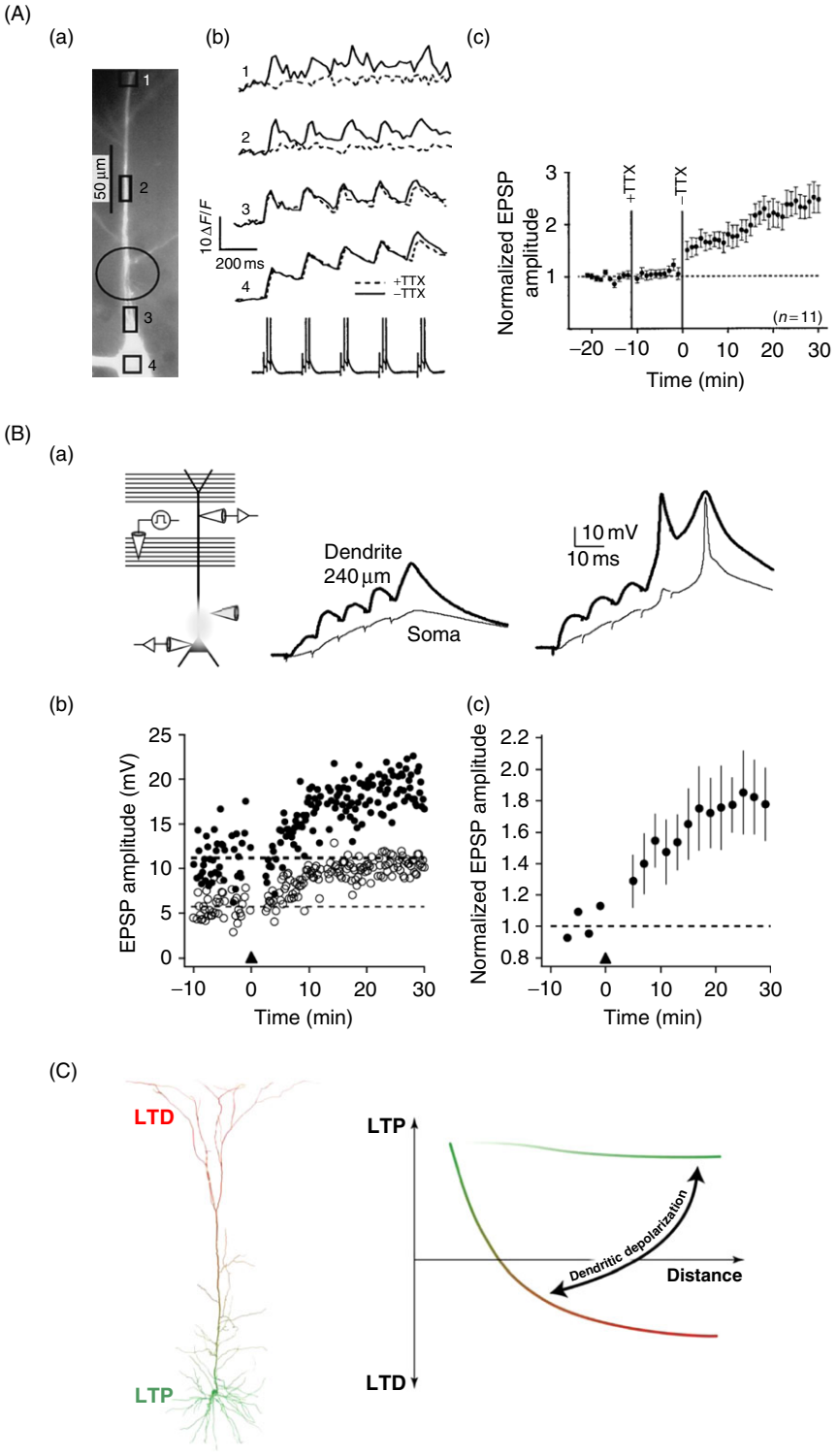
the result is often a burst of action potentials in the axon, which is likely to more powerfully promote the release of neurotransmitter from the axon terminals, compared to a single action potential. In Purkinje cells, however, the main function of the burst of action potentials may be the long pause in firing that follows the burst. Purkinje cells fire spontaneously and release the inhibitory transmitter GABA, so the pause in firing that follows the Ca^{2+} spike-induced burst may ultimately serve to activate the postsynaptic neurons in the deep cerebellar nucleus by releasing them from tonic inhibition.

A second possible function of dendritic excitability, besides promoting action potential firing in the axon, is to mediate neurotransmitter release from dendrites. The best-studied example of this is in the olfactory bulb, where reciprocal dendro-dendritic synapses are formed between mitral cells and granule cells. Both cell types have been shown to have excitable dendrites, which likely plays a key role in the activation of these synapses. Even in dendrites lacking structurally identifiable presynaptic specializations, dendrites can release substances that mediate local communication. For example, both pyramidal neurons and Purkinje cells have been shown to release endocannabinoids from their dendrites in response to dendritic depolarization. Events mediated by voltage-gated channels in dendrites (namely backpropagating action potentials and dendritic spikes) are likely to promote the release of these and other substances from dendrites. Notably, dendritic release may also be an outcome of the coincidence detection operations described above.

A third possible function of dendritic excitability is the induction or regulation of changes in synaptic strength such as Hebbian synaptic plasticity (Fig. 11.8). What kinds of dendritic events could provide the depolarization necessary for Hebbian plasticity? Backpropagating action potentials and dendritic spikes are both good candidates. For example, the backpropagating action potential has been proposed as an important signal for the induction of spike-timing-dependent synaptic potentiation and depression, because the activation of Ca_v channels (backpropagating action potential alone) and the relief of Mg^{2+} block of NMDA receptors (backpropagating action potential together with the binding of presynaptically released glutamate to the receptor) are sensitive to the relative timing of presynaptic and postsynaptic activation. Other forms of synaptic plasticity can occur in the absence of backpropagating action potentials. For example, dendritically initiated spikes have been shown to promote these forms of long-term potentiation (LTP) of excitatory synaptic strength even in the absence of axonal action potential firing. For some forms of plasticity, events mediated by dendritic excitability (backpropagating action potentials and dendritic spikes) are a requirement for the induction of synaptic plasticity. Therefore, these events are likely to serve as detectors of postsynaptic coincidence detection, thus mediating the property of synaptic cooperativity in LTP—that is, the requirement for multiple synaptic inputs to cooperate to produce the depolarization necessary for the induction of Hebbian LTP.

Although not as extensively studied, dendritic excitability may also contribute to other forms of plasticity, such as non-Hebbian synaptic plasticity, plasticity of inhibitory synapses, and plasticity of intrinsic membrane properties, including modulation of the very channels responsible for dendritic excitability. The latter mechanism could either work in a homeostatic manner or could enhance the probability of future dendritic spiking or synaptic plasticity. Finally, it should be noted that other processes might also regulate dendritic excitability. For example, modulatory neurotransmitters can regulate the activity of dendritic Na_v, Ca_v, and K_v channels, thus changing the excitable properties of the dendrite. This positions dendritic excitability as a mechanism that is both a regulator and a target of various forms of plasticity.

FIGURE 11.8 Dendritic excitability and synaptic plasticity. (A) Blocking backpropagating action potentials with dendritic application of TTX prevents the induction of LTP by pairing EPSPs with action potential firing. (a) CA1 pyramidal neuron filled with a Ca^{2+}-sensitive dye showing approximate location of TTX application (circle), strategically positioned to block action potential backpropagation without blocking action potential firing at the soma or synaptic transmission at more distal dendritic locations. (b) Ca^{2+}-sensitive fluorescence changes during pairing of EPSPs and action potentials (evoked by somatic current injection) in control (solid lines) and during application of TTX to the proximal dendrite (dashed lines). Numbered traces correspond to the locations of the numbered boxes in (a). Bottom trace is somatically recorded membrane potential during pairing. (c) Group data showing failure to induce LTP during dendritic TTX application, but not in the absence of TTX application. (B) Stronger stimulation can induce LTP in CA1 pyramidal neurons even when action potential backpropagation is blocked. (a) Paired somatic and dendritic recording during activation of presynaptic axons (no pairing with somatic current injection) and application of TTX to the soma and proximal dendrites to prevent action potential generation and backpropagation from the axon. Synaptic stimulation is strong enough to evoke dendritic spikes, which are attenuated at the soma. (b) Theta-burst synaptic stimulation induces LTP, which is observed in both the soma and the dendrite. (c) Group data showing LTP induction of somatic EPSP in experiments with backpropagating action potentials blocked. (C) Pairing of EPSPs and backpropagating action potentials in neocortical layer 5 pyramidal neurons induces either LTP or LTD, depending on the amplitude of the backpropagating action potentials. In proximal dendrites, where backpropagating action potentials are relatively large, LTP is induced. In distal dendrites, where backpropagating action potentials are smaller, LTD is induced. Depolarization of the distal dendrite amplifies backpropagating action potentials, however, resulting in induction of LTP instead of LTD. *Sources: (A) Magee and Johnston, 1997; (B) Golding et al., 2002; (C) Sjöström and Häusser, 2006.*



Dendritic Inhibition and Dendritic Excitability

One of the great mysteries of synaptic integration is why there are so many different types of inhibitory interneurons. Using classification mechanisms that depend on morphology, axon projections, physiological properties, and molecular markers, a great diversity of inhibitory cell types has been described. For example, more than 20 different types of inhibitory interneurons have been described in the CA1 region of the hippocampus alone. Interestingly, more than half of these types of interneurons

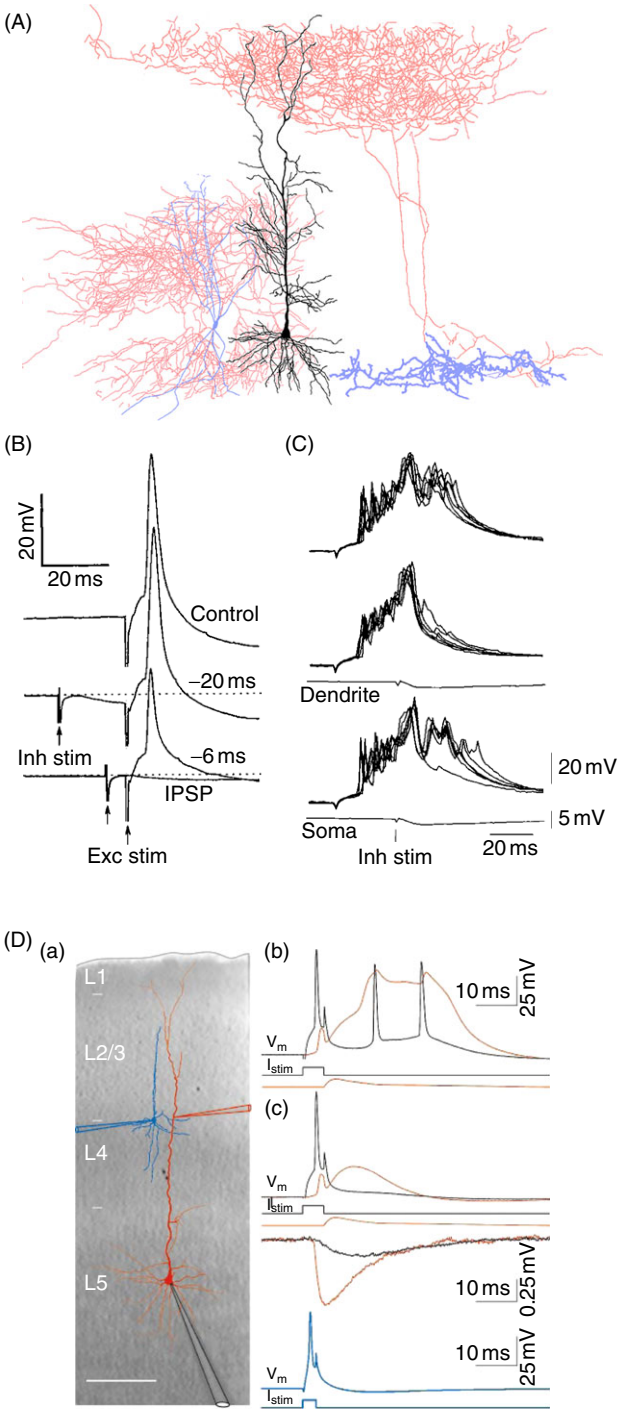


FIGURE 11.9 Functions of dendritic inhibition. (A) Compilation of inhibitory interneurons (dendrites blue, axon red) superimposed on a CA1 pyramidal neuron (black) to illustrate dendritic inhibition. (B) Effects of inhibition on backpropagating action potential in a CA1 pyramidal neuron. (C) Spikes evoked by current injection in a recording from a CA1 pyramidal neuron (top) are inhibited effectively by dendritic inhibition (middle), but not somatic inhibition (bottom). (D) (a) Simultaneous recording from a pyramidal neuron (red, soma and dendrite) and a neighboring inhibitory interneuron (blue). Scale bar is 200 µm. (b) BAC spike evoked by simultaneous somatic current injection to evoke an action potential and synaptic stimulation in layer 1 (somatic recording black; dendritic recording red). (c) BAC spike is blocked by activation of an action potential in the inhibitory interneuron (blue). IPSPs in the pyramidal neuron are also shown (i.e., with no current injection in pyramidal action potential or activation of layer 1 synapses). *Sources: (A) N. Spruston, McQuiston and Madison, 1999, Halasy et al., 1996; (B) Tsubokawa and Ross, 1996; (C) Miles et al., 1996; (D) Larkum et al., 1999.*

have axons that appear to innervate pyramidal neuron dendrites (Fig. 11.9A). What is the function of dendritic inhibition and why are the neurons that mediate it so diverse? At the moment, we have only limited insight into this problem, owing to the limited experimental data.

One of the most obvious functions of dendritic inhibition would be to selectively reduce the impact of excitatory synaptic inputs in the same dendritic compartment. For example, an interneuron that innervated the apical tuft of a pyramidal neuron might powerfully inhibit excitatory inputs to the tuft, while having little or no effect on excitatory synapses on basal dendrites. That said, the excitatory inputs that are affected by dendritic inhibition need not overlap directly with the inhibitory inputs. As noted above, for inhibition to be most effective, it should be on the path from the site of excitatory synapses to the axon.

Another likely function of dendritic inhibition is to regulate events mediated by dendritic excitability, such as backpropagating action potentials and dendritic spikes. Indeed, there is experimental evidence that these events can be attenuated by dendritic inhibition (Fig. 11.9B and 11.9C). Understanding the roles of different types of interneurons in this process is complex, however, because different varieties of dendritic spikes can be initiated in different dendritic compartments. To fully appreciate the role of dendritic inhibition, it will be necessary to understand not only the postsynaptic actions of each type of interneuron but also to understand when and why each interneuron type is active. For example, one study showed that feedback inhibition from one type of interneuron targets the proximal apical dendrites. During sustained firing, this form of dendritic inhibition outlasted somatic inhibition provided by a different type of interneuron. This cell-type-specific form of inhibition results from a combination of the intrinsic and synaptic properties of each type of interneuron. The ultimate goal is to understand when and why each cell type is activated in the intact brain and how this activity affects circuit function to mediate behavior and cognition.

STRUCTURE AND FUNCTION OF DENDRITIC SPINES

The dendrites of many neurons are covered with thousands of dendritic spines—small protuberances that form the primary sites of contact for excitatory synapses (and some inhibitory synapses) in many types of neurons. Ever since their discovery by Ramon y Cajal, the

function of spines has been a mystery. Today, we know a great deal more about the electrical and chemical properties of spines, but the reason for their existence remains elusive. We know for sure that dendritic spines are not essential for the basic function of excitatory synapses, since many neurons—most notably inhibitory interneurons—receive excitatory synapses directly onto their dendrites and lack dendritic spines. Even in some spiny neurons, some excitatory synapses are formed directly onto the dendrite, raising the possibility that excitatory synapses on spines and dendrites may have distinct functional properties. Certainly, the abundance of spines in the nervous system suggests strongly that they serve an important function.

Structure of Dendritic Spines

The typical dendritic spine is often described as having a "mushroom" shape, consisting of a narrow spine neck and a larger spine head (about one micron diameter). However, the structure of dendritic spines can be quite varied, even for spines on a single neuron (Fig. 11.10). Some spines have particularly long necks, which tends to correlate with smaller spine heads, while other spines have short, stubby necks and an unclear distinction between the neck and the head. In electron micrographs, spines are characterized by the presence of smooth endoplasmic reticulum and an actin cytoskeleton, as well as by the absence of microtubules, ribosomes, and mitochondria. In some cases a particularly extensive endoplasmic reticulum forms a structure called a "spine apparatus." One of the most distinctive features of dendritic spines is the postsynaptic density (PSD), which is an electron-dense region opposite the presynaptic terminal, thought to contain the postsynaptic machinery for synaptic function. Typically, PSDs are largest in mushroom-shaped spines and smaller in thin or stubby spines. As noted above, some spines have particularly large PSDs that are interrupted by a perforation. Strong correlations have been noted among the size of the spine head, the size of the PSD, and the number of AMPA-type glutamate receptors, indicating that there is a relationship between spine size and synaptic strength.

It has been suggested, and recently confirmed by experimental observations, that structural changes in dendritic spines can occur together with synaptic plasticity. This is a seemingly simple idea—stronger synapses need bigger spines, so growth of the spine is required for synaptic potentiation (e.g., LTP)—but complexities and open questions lurk beneath the surface of this idea. Can changes in spine structure alone mediate changes in synaptic strength, or are other molecular changes required as well? Are thin spines the initial precursors of mushroom spines and stubby spines the remnants of retracting spines, or does spine diversity reflect a diverse population of stable synapses? Indeed, conversion between the different types of spines has been observed *in vitro* and *in vivo*. A consensus from the *in vivo* studies is that some spines change over time, while other spines are extremely stable, reflecting the need for a balance between plasticity and stability in neural circuits.

Chemical and Electrical Compartmentalization in Spines

Conjecture regarding the function of dendritic spines has focused on two possibilities. The first is that spines function as a chemical compartment distinct from the dendritic shaft (Fig. 11.11A and 11.11B). This idea is suggested by the narrow constriction of the spine neck, which has been shown experimentally to limit the diffusion of substances out of the spine head. Even small molecules such as Ca^{2+} do not diffuse freely out of spines. As a result, during activation of NMDA receptors on spines, the intracellular Ca^{2+} concentration in the spine head can far exceed that of the dendritic shaft directly adjacent to the spine. Such observations, which have been greatly facilitated by the development of two-photon microscopy (Box 11.3), have been used to quantify the diffusion of small fluorescent molecules into and out of the spine. Two-photon, as well as conventional, light microscopy has also been used to study other chemical processes in spines. For example, the endoplasmic reticulum in spines has been shown to provide a source for Ca^{2+} release into the spine head, buffering of spine Ca^{2+} concentration, as well as a source for membrane proteins such as AMPA receptors, which are inserted into the spine-head membrane during LTP. Indeed, chemical compartmentalization in the spine head is likely to be crucially important for activity-dependent plasticity, such as LTP, which may depend initially on Ca^{2+} and subsequently on an orchestrated series of biochemical and cell biological processes in the spine head.

The other major concept central to the function of dendritic spines is electrical compartmentalization (Fig. 11.11B and 11.11C). Because they are so small, the input resistance of spines is higher than that of most dendrites. The value of spine input resistance depends primarily on the resistance of the spine neck, because the parallel resistance of the spine head is immeasurably large (on the order of thousands of gigaohms). One way of thinking about this is that there is very little membrane (and therefore very few channels) through which charge can flow out of the spine head. The spine neck, therefore, although thin, is by comparison a lower-resistance path out of the spine (and into it as well). Estimates of spine neck resistance are highly varied, ranging from approximately 50 megaohms (similar to the input impedance of large dendrites) to more than a gigaohm (similar to the

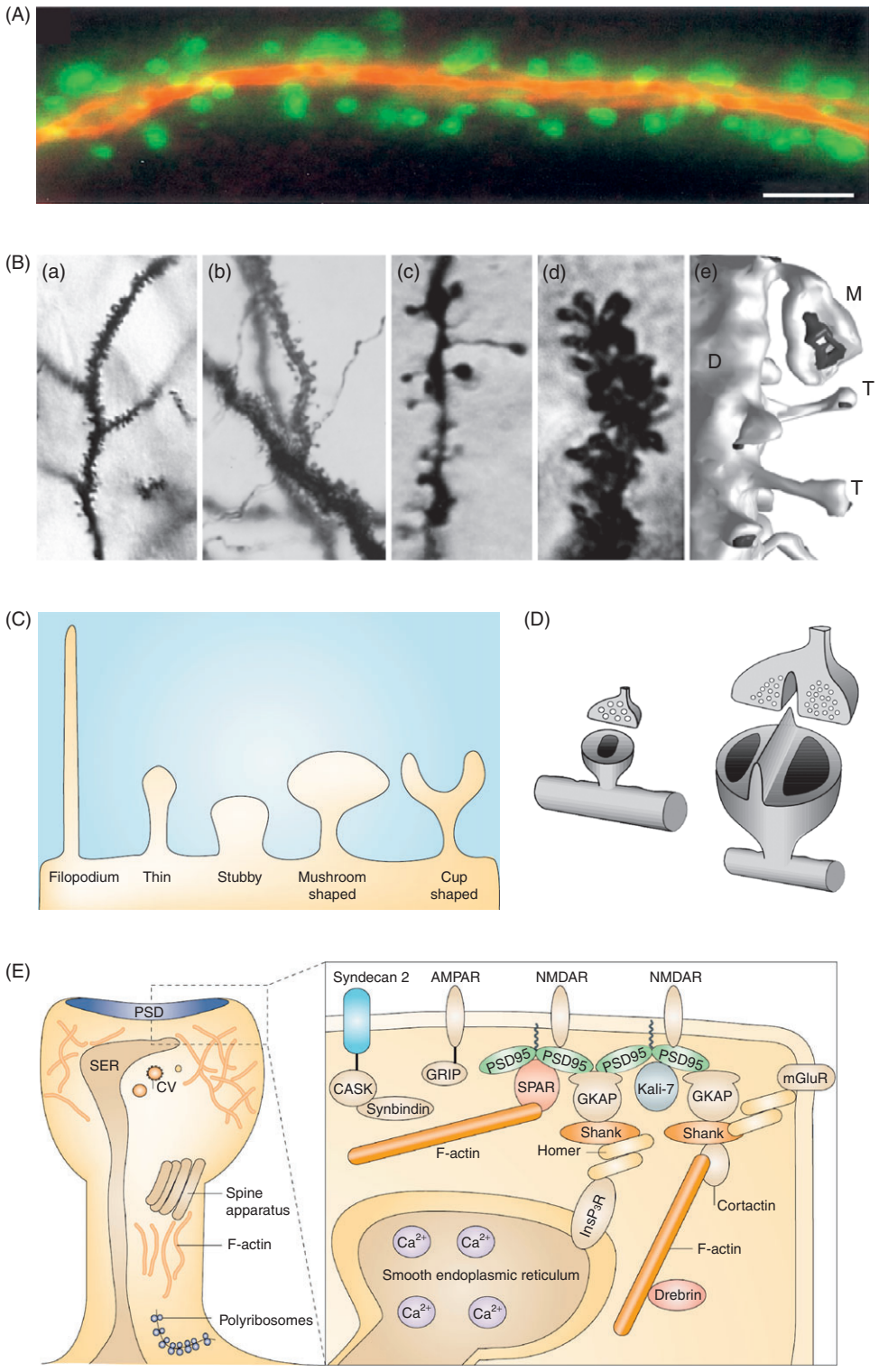


FIGURE 11.10 Spine diversity. (A) Dendritic spines expressing GFP-tagged actin (green) decorate a MAP2-labeled dendrite (red) of a cultured hippocampal neuron. Scale bar is 5 μm. (B) Dendritic spines at increasing magnification from: (a) neocortical pyramidal neuron, (b) hippocampal CA1 pyramidal neuron, (c) neocortical pyramidal neuron, (d) cerebellar Purkinje cell, (e) CA1 pyramidal neuron showing mushroom (M) and thin (t) pedunculated spines. (a–d) are light microscopy and (e) is an EM reconstruction. (C) Cartoons of different spine shapes. (D) Cartoons of small and large axo-spinous synapses. Note the "perforation" of the PSD in the large synapse. In CA1 pyramidal neurons, approximately 20% of synapses have perforated PSDs. These PSDs also have a higher density of AMPA receptors than nonperforated PSDs, suggesting that they are especially powerful synapses. (E) Cartoon illustrating the most abundant organelles and molecules in spines, including: smooth endoplasmic reticulum (SER), coated vesicles (CV), F-actin, and numerous other molecules associated with the postsynaptic density (PSD). *Sources: (A) Matus, 2000; (B) Fiala et al., 2008; (C) Hering and Sheng, 2001; (D) Geinisman, 2000; (E) Hering and Sheng, 2001.*

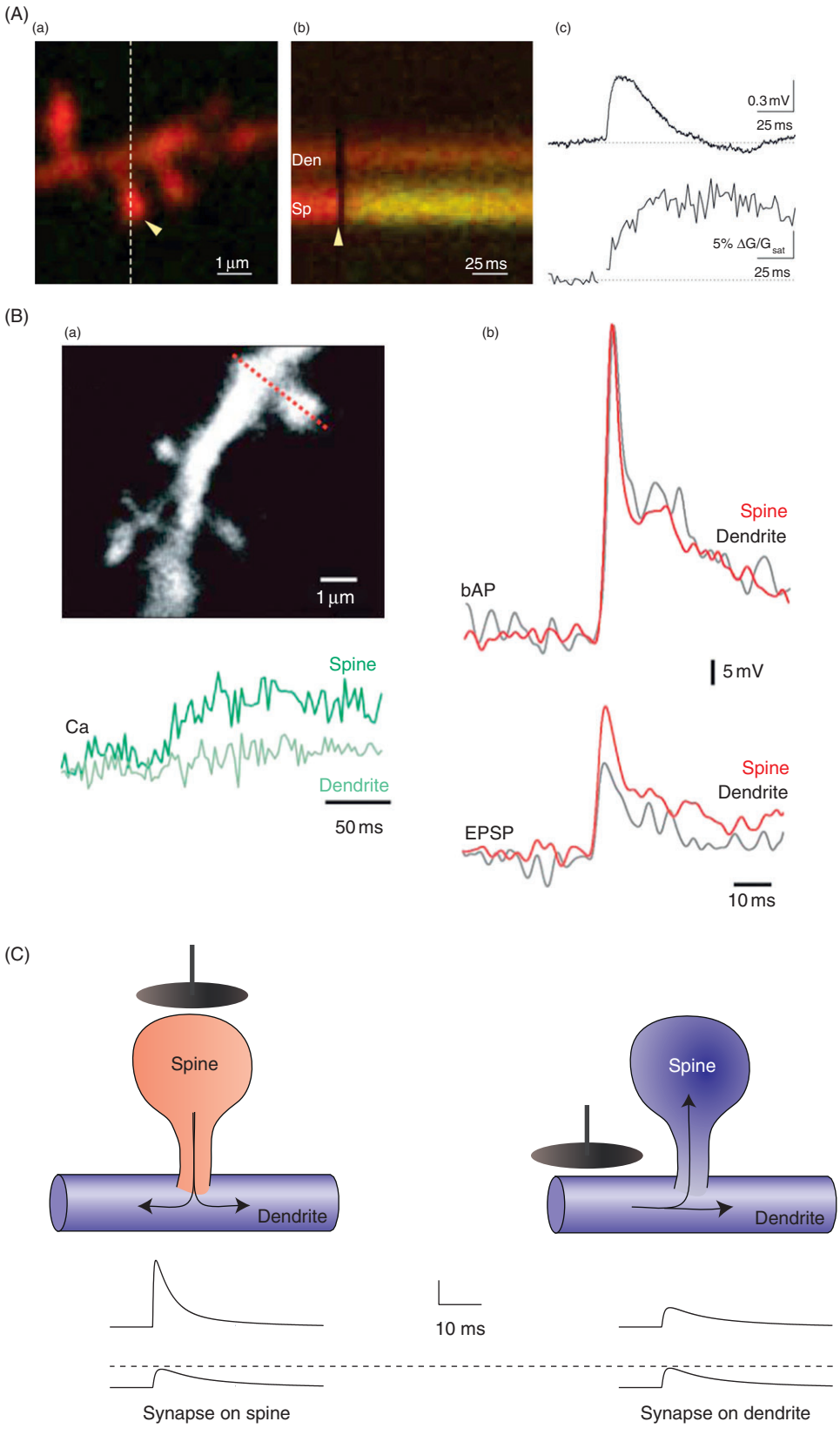


FIGURE 11.11 Chemical and electrical compartmentalization in spines. (A) (a) Image of a dendrite from a CA1 pyramidal neuron filled with a Ca^{2+}-insensitive dye (red) and a Ca^{2+}-sensitive dye (green). (b) Time series showing fluorescence changes along the line scan covering the spine (Sp) and the dendrite (Den). The yellow arrowhead indicates the position of a brief laser pulse to uncage MNI-glutamate at the position indicated in (a). (c) Somatically recorded EPSP (top) and Ca^{2+}-mediated fluorescence increase (bottom) in the spine in response to uncaging. (B) (a) Ca^{2+} transients (green traces below image) restricted to the spine in layer 5 pyramidal neuron. (b) Voltage-sensitive fluorescence changes in the spine (red traces) and dendrite (black traces) in response to a backpropagating action potential (bAP) and an EPSP. (C) Schematic diagram of the responses to a synapse on the spine head (left) and on the dendrite (right). Red indicates larger EPSP than blue. In B and C, note the attenuation from spine to dendrite, but not from dendrite to spine. *Sources: (A) Bloodgood and Sabatini, 2007; (B) Palmer and Stuart, 2009; (C) N. Spruston.*

input impedance of very small, terminal dendrites). This large range may reflect the range of spine-neck geometries: short, stubby spines likely have a relatively low resistance, while long, thin spines should have much higher resistances. Technical difficulties associated with estimating spine neck resistance may also contribute to this large range.

The spine can be considered electrically as a very high-resistance spine head connected to a relatively low-resistance dendrite via an intermediate-resistance spine neck. This electrical circuit consists of multiple impedance mismatches, which means that voltage drops are highly asymmetrical: voltage drops considerably from the spine head to the dendritic shaft, but it should hardly change at all when spreading in the other direction. Functionally, this means that when an excitatory synapse on a spine head is activated, the voltage in the spine head should be considerably larger than the voltage in the dendrite to which it is attached. On the other hand, voltages in the dendrite are expected to transmit reliably (i.e., without significant attenuation) into the spine head. This is biophysically similar to the situation described earlier for the dendrites: Dendritic spikes propagate poorly in the forward direction (from smaller dendrites into larger dendrites), but backpropagating action potentials propagate more reliably in the reverse direction (from larger dendrites into smaller dendrites).

The functional consequences of the electrical properties of spines are not known with certainty, but a few clear predictions can be made based on the electrical model described above. First, even though a single synapse produces very small EPSPs in the soma, the EPSP in the spine head is expected to be quite large, which could lead to activation of voltage-gated channels or relieve the Mg^{2+} block at activated NMDA receptors in the spine head. Indeed, Ca^{2+} imaging in spines has revealed Ca^{2+} entry through Ca_v channels and NMDA receptors, even in response to synaptic activation of a single spine. Second, the large voltage drop from the spine to the dendrite will limit the extent to which each synapse can influence its neighbors. For example, if the voltage drop were 10-fold, a 10 mV EPSP in the spine would only produce a 1 mV depolarization of neighboring spines. Third, voltage in the dendrite transmits reliably into spines. This makes it possible for a group of spines to influence one another, as well as for backpropagating action potentials or dendritic spikes to change the membrane potential in the spine. Indeed, this prediction has been validated by observations that backpropagating action potentials can readily activate Ca_v channels and relieve Mg^{2+} block of NMDA receptors in spines, a property that is crucially important for the ability of dendritic excitability to influence synaptic plasticity.

Another important question regarding the electrical properties of spines that has been considered ever since they were first discovered by Ramon y Cajal is whether changes in the spine neck can alter the coupling between the spine and the dendrite to change synaptic strength. Although this idea is appealing, the answer is probably "no." The spine neck is so short that it has negligible membrane resistance and capacitance. Thus, the amount of charge that is lost from the neuron as charge flows from the spine head to the dendrite is expected to be minuscule—certainly not large enough to significantly alter the size of the EPSP. This prediction is expected even for the highest-resistance spine necks, so lowering the neck resistance is not expected to increase the amount of charge reaching the dendrite. Changes in the spine neck can, however, have other effects. For example, lowering spine neck resistance would be expected to reduce the spine-head EPSP produced by activating the synapse on the spine, thus reducing the activation of voltage-gated channels in the spine head. Attempts to measure such effects, although difficult, have suggested that these effects are at best very small. Similarly, because of the high input impedance of the spine head, changes in the neck resistance are not likely to significantly affect the propagation of voltages from the dendrite into the spine.

Together, the available data and computer models suggest that the main electrical effect of spines is to amplify the local voltage at the synapse, perhaps allowing for more effective activation of Ca_v channels and relief of Mg^{2+} block at NMDA receptors. The resultant Ca^{2+} entry will elevate the Ca^{2+} concentration in the spine, influencing Ca^{2+}-sensitive biochemical processes in the spine head. Any resultant changes in spine structure are expected to have only minor effects on the electrical properties of spines, but the chemical compartmentalization provided by spines may facilitate biochemical and cell biological processes that influence synaptic strength such as addition of AMPA receptors and restructuring of the PSD.

COMPUTATIONS PERFORMED BY DENDRITES

Detailed analysis of synaptic integration in dendrites has been motivated by the implicit assumption that the ability of neural circuits to process and store information is dependent on the processing power of single neurons. This seems like a reasonable assumption when one considers the complexity of dendrites and their variability between different neuronal types. Ultimately, testing this assumption will require that dendrites be manipulated—likely through genetic means—to investigate how changes in their structure or ion channel expression affect neural circuit function and behavior. This is a great challenge, but one that is beginning to be possible owing to the development of sophisticated molecular genetic techniques. In the meantime, work has proceeded, using a

combination of anatomical, physiological, and computer-modeling approaches focused on developing working hypotheses about how the properties of dendrites may be used to perform neuronal "computations."

Boolean Operations and Dendritic Arithmetic

The properties of dendrites described in the previous sections allow dendrites to perform some basic Boolean operations such as AND, OR, or AND-NOT operations, thus potentially imparting onto neurons the ability to serve as basic computational elements.

One way that neurons could implement an AND operation is via the activation of two different types of presynaptic inputs targeting different dendritic domains (Fig. 11.12A)—in other words, simple coincidence detection. For example, CA1 pyramidal neurons receive inputs from CA3 (proximal inputs) and entorhinal cortex, (EC, distal inputs). These two inputs could be used to perform an AND operation (CA3 AND EC input causes CA1 spiking) or an OR operation (CA3 OR EC input causes CA1 spiking). It seems more likely to implement the AND operation, however, because EC inputs are quite far from the soma, with recordings and modeling studies suggesting that EC input is unlikely to trigger axonal action potential firing on its own. In this example, the AND operation may arise from the ability of the two inputs to interact nonlinearly in the dendrites by virtue of voltage-gated channels giving rise to dendritic spikes. Intriguingly, the observation that some

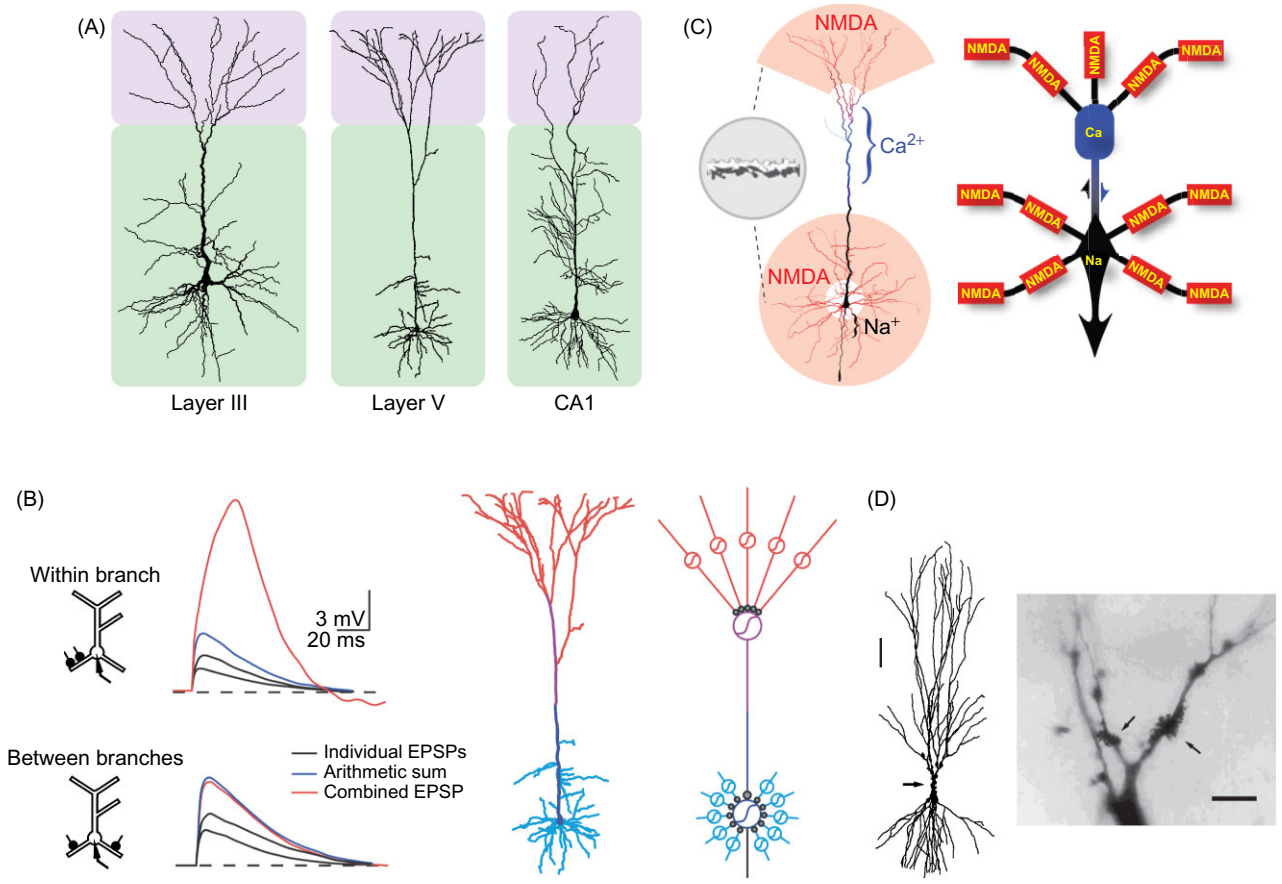


FIGURE 11.12 Compartmentalization of different dendritic inputs is important for dendritic computation. (A) Schematic diagram illustrating that the apical tuft (purple) and proximal apical/basal dendrites (green) receive separate synaptic inputs in a variety of types of pyramidal cells. (B) Summation of inputs onto a single dendritic branch (but not between branches) leads to a dendritic spike that remains below threshold for an action potential at the soma. It has therefore been proposed that each proximal apical and basal dendritic branch (blue) acts like a computational unit where synaptic inputs are summed (represented by sigmoidal functions in small circles) and multiple branches need to be activated to trigger an action potential (represented by sigmoidal function in large circles). Distal apical dendrites (red) may similarly sum inputs, which are further summed in a distal dendritic spike-initiation zone. (C) This model proposed in B has been further validated by experiments demonstrating summation of NMDA spikes in apical tuft dendrites to activate a distal Ca^{2+} spike-initiation zone. (D) CA3 pyramidal neurons receive strong excitatory synaptic input via the mossy fibers from dentate granule cells. These synapses contact elaborate dendritic spines (arrows) called thorny excrescences. Because of the large size of each mossy fiber synapse, they have been referred to as "detonator" synapses. Scale bars: 50 μm (left), 25 μm (right). *Sources: (A) Spruston, 2008; (B) Polsky et al., 2004, Spruston and Kath, 2004; (C) Larkum et al., 2009; (D) Gonzales et al., 2001.*

excitatory inputs selectively target distal dendrites and others target proximally is a common theme in many kinds of neurons (e.g., pyramidal neurons, dentate granule cells), suggesting that AND operations may be commonly implemented in this way.

Other mechanisms can also lead to AND operations in dendrites. For example, as discussed above, spikes generated in dendrites do not propagate reliably to the soma. Thus, dendritic spikes are often observed even in the absence of an axonal action potential. One idea that has been proposed is that pyramidal neurons might fire action potentials only when two or more dendritic branches are depolarized to threshold for a dendritic spike (Fig. 11.12B). For more distal synaptic inputs, such as those onto the apical tuft of layer 5 pyramidal neurons, NMDA spikes in multiple dendritic branches may be required in order to trigger an action potential in the axon (Fig. 11.12C). Thus, neurons can implement an AND operation using multiple dendritic branches. Another mechanism is the BAC spike observed in layer 5 pyramidal neurons (discussed above). BAC spikes implement an AND operation using excitatory input to the apical tuft combined with a backpropagating action potential (e.g., evoked by strong excitatory input to the basal dendrites).

Although the simplest possible outcome of a dendritic AND operation is generation of an action potential in the axon, other possible readouts may occur as well. In the case of a distal dendritic Ca^{2+} spike or BAC spike, the readout may be a burst of action potentials in the axon, or a large dendritic event that can influence synaptic plasticity even in the absence of an axonal spike. In Purkinje cells, activation of the climbing fiber alone always leads to a burst of action potentials, but when combined with parallel fiber activation, the AND operation may lead to more spikes in the burst, as well as plasticity at the activated parallel fiber inputs. Even individual dendritic spines can perform AND operations, as activation of glutamate receptors combined with a backpropagating action potential can lead to Ca^{2+} influx through NMDA receptors, thus influencing biochemical reactions in the spine, including those that can lead to synaptic plasticity.

Synaptic integration in dendrites may also give rise to OR operations. For example, hippocampal CA3 pyramidal neurons receive input from granule cells in the dentate gyrus via mossy fibers. The mossy fiber axons form large, specialized excitatory synapses on the main apical dendrite, very close to the soma (Fig. 11.12D). These synapses are sometimes called "detonator" synapses, because they produce large EPSPs that can cause the CA3 neuron to fire action potentials. On the other hand, CA3 neurons also form extensive collateral connections among themselves. Many of these synapses are also quite close to the soma, on basal and proximal apical dendrites, so they can also cause CA3 neurons to fire. Thus, CA3 neurons may respond to mossy fiber OR collateral input alone. CA3 neurons also receive excitatory input from EC onto the distal apical dendrites, however. Thus, like CA1 neurons, CA3 neurons could implement an AND operation as well (CA3 collateral AND EC input cause CA3 spiking). This anatomical arrangement suggests that neurons could implement both AND and OR operations, depending on which inputs are active in the dendritic tree.

Neurons can also use inhibition to implement AND-NOT Boolean operations. In this case, inhibition is viewed as vetoing the excitatory input. As discussed in previous sections, however, the effectiveness of inhibition depends not only on its strength but also on its location relative to active excitatory inputs. Thus, inhibition may perform a global AND-NOT operation if powerful inhibition is targeted to the soma or axon, or it may perform a more local AND-NOT operation if the inhibition is targeted to a particular dendritic domain. Many cortical pyramidal neurons receive feedback inhibition to the distal apical tuft, suggestive of a veto (NOT) of distal excitatory inputs.

Another mechanism by which neurons might implement Boolean operations is by using inputs from neuromodulatory systems. For example, a combination of excitatory synaptic input AND cholinergic input could be a condition for action potential firing, while excitatory input AND-NOT serotonergic input might be required. In these hypothetical scenarios, the cholinergic and serotonergic inputs could be excitatory and inhibitory, respectively, by acting on synaptic targets (e.g., enhancing or inhibiting glutamate release) or nonsynaptic targets (e.g., enhancing or inhibiting voltage-gated channels in the axon, soma, or dendrites).

Although these Boolean operations are relatively simple, very complex operations can be performed using them, as is the case for modern computers. Neural circuits can do this using large numbers of neurons, but as noted above, it may also be possible for one neuron to perform multiple (perhaps even many) Boolean operations. This can happen either separately or simultaneously. CA3 neurons may perform separate AND or OR operations, depending on the inputs they receive in any particular behavioral condition. On the other hand, action potential firing in some cases might represent a complex condition resulting from multiple, simultaneous operations, such as the following:

Distal excitation AND-NOT distal inhibition
AND
Proximal excitation AND dopaminergic modulation
AND-NOT
Axo-somatic inhibition

In this hypothetical scenario, some operations are on the same line to indicate local specificity of the effects,

while others are on separate lines to indicate the more global nature of the operation. As illustrated by this example, single neurons may implement complex operations that have traditionally been assigned to neural networks consisting of many neurons, each with much simpler properties. Thus, the presence of excitable dendrites that receive targeted inputs from distinct populations of excitatory, inhibitory, and modulatory synapses may confer complex computational power to single neurons.

In addition to performing Boolean operations, dendrites have also been described as performing arithmetic operations. The simplest example of this is summation. If a number of excitatory synaptic inputs are summed (spatially or temporally) to reach action potential threshold, this can be thought of as a simple addition. On the other hand, inhibitory inputs can reduce the depolarization from the summed excitatory inputs by providing a hyperpolarizing influence, which can be thought of as subtraction.

More complicated arithmetic—namely multiplication and division—can also be implemented in dendrites. For example, the AND operations described above can be thought of as multiplication because if either input is zero, the output is zero. On the other hand, shunting inhibition can be thought of as division because it reduces depolarization in proportion to the strength of the excitatory input (larger excitatory inputs produce larger driving forces for current flowing through the inhibitory shunt).

Dendrites and Neuronal Response Properties

Moving a step beyond the Boolean and arithmetic operations performed by dendrites (and the biophysical mechanisms used to implement them) requires a link between the function of a neuron (i.e., its response properties) and the properties of its dendrites. Such a link has been made in several cases, but in general the mechanisms are not well established. Two examples stand out, however, as illustrations of how dendrites can contribute to determining response properties.

The first example is the general case of visual motion detection, which has been studied in a variety of cell types with motion-sensitive receptive fields. Although several mechanisms have been described, what they all have in common is sequentially activated synaptic inputs targeting different dendritic positions in spatial order. Even in passive dendrites, it has been shown that sequential activation of excitatory synapses from distal (first) to proximal (last) is more effective than activation of the same synapses in reverse order (Box 11.1). The reason is that distal inputs have slower kinetics when they arrive at the soma, so they sum most effectively with proximal inputs (with faster kinetics) that arrive later. This mechanism is enhanced in dendrites by NMDA receptors, but for a different reason: distal dendrites have higher input impedance than proximal dendrites, so they produce large depolarizations that more effectively remove Mg^{2+} block at NMDA receptors, both at the distal synapses and at proximal synapses that are activated later. When synapses are activated in the reverse order (proximal before distal), the large dendritic depolarization associated with activation of distal synapses occurs too late to effectively remove the Mg^{2+} block at the proximal synapses activated earlier.

A second example of a well-understood mechanism of dendritic integration underlying a response property comes from the auditory system, where neurons involved in sound localization are responsive to interaural timing differences. These neurons respond best to coincident input from the two ears, with timing differences encoded by "delay lines" that allow for coincident arrival of inputs to the two ears even when sound arrives at the two ears with a slight delay. To respond optimally to two coincident inputs with very rigid timing requirements (microseconds), the EPSPs have to decay very rapidly, which requires very leaky dendrites. If the two inputs are located on the same dendrite, they will shunt each other (i.e., reduced driving force at one synapse resulting from depolarization via the other), so to effectively summate, the two inputs must be formed onto two different dendrites. Because these two dendrites are on opposite sides of the soma, the EPSP from one dendrite has little effect on the EPSP from the other dendrite, thus limiting the shunting effect. By contrast, a higher-intensity sound arriving at one ear would not be an effective stimulus for the neuron to fire because of the suppressive nonlinearity (shunting) caused by activating more synapses on the same dendrite.

The common denominator in these mechanisms is that in order for them to work, synaptic inputs have to be targeted to precise locations. It is not good enough for an axon carrying particular information (i.e., sound from the left ear) to find the correct postsynaptic neuron; it must also find the correct dendrite (i.e., left vs. right) on that neuron. Similarly, in pyramidal neurons, inputs clustered onto the same dendritic branch trigger dendritic spikes more readily than inputs that are distributed across multiple branches, suggesting that even more precise targeting may be required. Taking this a step further, clustering of inputs alone may not even be precise enough. For example, directional selectivity may require that certain synapses must find not only the correct dendrite but also the correct location on the dendrite (i.e., distal vs. proximal). The degree to which neural circuits are wired with this degree of precision is unknown. For example, do multiple axons carrying common information (e.g., a particular facial feature) selectively target the same dendrites or dendritic

compartments? Questions like these must be answered before we can truly understand how information is processed in dendrites.

SUMMARY

Neuronal dendrites are very complicated. This is obvious from a glance at their structure in the light microscope but is becoming even more apparent as the molecular composition and functional properties of different dendrites are investigated. In addition to the channels that give them a resting potential and allow them to engage in synaptic communication, they contain a host of voltage-gated channels expressed in exquisite patterns, differentially distributed across the dendrites of different neurons. In addition, synapses carrying information from different parts of the brain selectively target different dendritic domains. We have only begun to scratch the surface of the complexity that exists in neuronal dendrites, and there are many important questions left to answer, spanning multiple levels of organization and complexity. For example, at the cellular level: why are there at least 50 times more HCN channels in distal dendrites than in proximal dendrites of many pyramidal neurons? At the synaptic and circuit levels: why are there more than 20 different types of inhibitory interneurons in the CA1 region of the hippocampus alone, more than half of which target specific dendritic compartments? How precise is the specificity of this dendritic targeting, not only for inhibitory synapses but also for excitatory synapses? At the computational level: how do these cellular properties and synaptic circuitry come together with the structure of the dendritic tree to implement sophisticated cellular computation? At the systems level: why are distal and proximal synapses differentially modulated by modulatory neurotransmitters such as acetylcholine? How does such modulation alter the computations performed in dendrites as neurotransmitter levels change during various behavioral states? These and many other questions need to be answered, likely leading to even more questions. Elucidating the details of how synaptic integration works in elaborately branching dendritic trees will be critical for determining not only how neurons are built, but for what purpose. Ultimately, this challenge lies at the core of understanding how neurons are used to build circuits controlling sensation and action, perception, and cognition.

References

Benardo, L. S., Masukawa, L. M., & Prince, D. A. (1982). Electrophysiology of isolated hippocampal pyramidal dendrites. *Journal of Neuroscience, 2,* 1614–1622.

Bloodgood, B. L., & Sabatini, B. L. (2007). Ca(2+) signaling in dendritic spines. *Current Opinion in Neurobiology, 17,* 345–351.

Branco, T., Clark, B. A., & Häusser, M. (2010). Dendritic discrimination of temporal input sequences in cortical neurons. *Science, 329,* 1671–1675.

Cole, K., & Hodgkin, A. (1939). Membrane and protoplasm resistance in the squid giant axon. *The Journal of General Physiology, 22,* 671–687.

Davis, L., & Lorente de Nó, R. (1947). Contribution to the mathematical theory of the electrotonus. *Studies from the Rockefeller institute for medical research. Reprints. Rockefeller Institute for Medical Research, 131,* 442–496.

Denk, W., & Svoboda, K. (1997). Photon upmanship: Why multiphoton imaging is more than a gimmick. *Neuron, 18,* 351–357.

Dodt, H. U., & Zieglgansberger, W. (1990). Visualizing unstained neurons in living brain slices by infrared DIC-videomicroscopy. *Brain Research, 537,* 333–336.

Edwards, F. A., Konnerth, A., Sakmann, B., & Takahashi, T. (1989). A thin slice preparation for patch clamp recordings from neurones of the mammalian central nervous system. Pflugers Archiv, *414,* 600–612.

Fiala, J. C., Spacek, J., & Harris, K. M. (2008). Chapter 1. Dendrite structure. In *Dendrites*. Oxford: Oxford University Press.

Geinisman, Y. (2000). Structural synaptic modifications associated with hippocampal LTP and behavioral learning. *Cerebral Cortex, 10,* 952–962.

Golding, N. L., Kath, W. L., & Spruston, N. (2001). Dichotomy of action-potential backpropagation in CA1 pyramidal neuron dendrites. *Journal of Neurophysiology, 86,* 2998–3010.

Golding, N. L., & Spruston, N. (1998). Dendritic sodium spikes are variable triggers of axonal action potentials in hippocampal CA1 pyramidal neurons. *Neuron, 21,* 1189–1200.

Golding, N. L., Staff, N. P., & Spruston, N. (2002). Dendritic spikes as a mechanism for cooperative long-term potentiation. *Nature, 418,* 326–331.

Goldstein, S. S., & Rall, W. (1974). Changes of action potential shape and velocity for changing core conductor geometry. *Biophysical Journal, 14,* 731–757.

Gonzales, R. B., DeLeon Galvan, C. J., Rangel, Y. M., & Claiborne, B. J. (2001). Distribution of thorny excrescences on CA3 pyramidal neurons in the rat hippocampus. *Journal of Comparative Neurology, 430,* 357–368.

Halasy, K., Buhl, E. H., Lörinczi, Z., Tamás, G., & Somogyi, P. (1996). Synaptic target selectivity and input of GABAergic basket and bistratified interneurons in the CA1 area of the rat hippocampus. *Hippocampus, 6,* 306–329.

Hamill, O. P., Marty, A., Neher, E., Sakmann, B., & Sigworth, F. J. (1981). Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pflugers Archiv, 391,* 85–100.

Häusser, M., Spruston, N., & Stuart, G. J. (2000). Diversity and dynamics of dendritic signaling. *Science, 290,* 739–744.

Hell, S. W. (2007). Far-field optical nanoscopy. *Science, 316,* 1153–1158.

Hering, H., & Sheng, M. (2001). Dendritic spines: Structure, dynamics and regulation. *Nature Reviews Neuroscience, 2,* 880–888.

Hodgkin, A., & Rushton, W. (1946). The electrical constants of a crustacean nerve fibre. *Proceedings of Royal Society of London, 133,* 444–479.

Larkum, M. E., Nevian, T., Sandler, M., Polsky, A., & Schiller, J. (2009). Synaptic integration in tuft dendrites of layer 5 pyramidal neurons: A new unifying principle. *Science, 325,* 756–760.

Larkum, M. E., Zhu, J. J., & Sakmann, B. (1999). A new cellular mechanism for coupling inputs arriving at different cortical layers. *Nature, 398,* 338–341.

Lee, A. K., Manns, I. D., Sakmann, B., & Brecht, M. (2006). Whole-cell recordings in freely moving rats. *Neuron, 51,* 399–407.

Llinás, R., & Sugimori, M. (1980). Electrophysiological properties of in vitro Purkinje cell dendrites in mammalian cerebellar slices. *Journal of Physiology* (London), *305,* 197–213.

Losonczy, A., & Magee, J. C. (2006). Integrative properties of radial oblique dendrites in hippocampal CA1 pyramidal neurons. *Neuron, 50,* 291–307.
Magee, J. C., & Johnston, D. (1997). A synaptically controlled, associative signal for Hebbian plasticity in hippocampal neurons. *Science, 275,* 209–213.
Matus, A. (2000). Actin-based plasticity in dendritic spines. *Science, 290,* 754–758.
McQuiston, A. R., & Madison, D. V. (1999). Nicotinic receptor activation excites distinct subtypes of interneurons in the rat hippocampus. *Journal of Neuroscience, 19,* 2887–2896.
Mel, B. W. (1994). Information processing in dendritic trees. *Neural computation, 6,* 1031–1085.
Micheva, K. D., & Smith, S. J. (2007). Array tomography: A new tool for imaging the molecular architecture and ultrastructure of neural circuits. *Neuron, 55,* 25–36.
Miles, R., Tóth, K., Gulyás, A. I., Hájos, N., & Freund, T. F. (1996). Differences between somatic and dendritic inhibition in the hippocampus. *Neuron, 16,* 815–823.
Miyawaki, A. (2011). Development of probes for cellular functions using fluorescent proteins and fluorescence resonance energy transfer. *Annual Review of Biochemistry, 280,* 357–373.
Nevian, T., Larkum, M. E., Polsky, A., & Schiller, J. (2007). Properties of basal dendrites of layer 5 pyramidal neurons: a direct patch-clamp recording study. *Nature Neuroscience, 10,* 206–214.
Palmer, L. M., & Stuart, G. J. (2009). Membrane potential changes in dendritic spines during action potentials and synaptic input. *Journal of Neuroscience, 29,* 6897–6903.
Polsky, A., Mel, B. W., & Schiller, J. (2004). Computational subunits in thin dendrites of pyramidal cells. *Nature Neuroscience, 7,* 621–627.
Rall, W. (1959). Branching dendritic trees and motoneuron membrane resistivity. *Experimental Neurology, 1,* 491–527.
Rall, W. (1964). Theoretical significance of dendritic trees for input-output relations. In *Neural Theory and Modeling*. Reiss, R.F. (Ed.). Palo Alto: Stanford University Press.
Scanziani, M., & Häusser, M. (2009) Electrophysiology in the age of light. *Nature, 461*, 930–939.
Schiller, J., Schiller, Y., Stuart, G. J., & Sakmann, B. (1997). Calcium action potentials restricted to distal apical dendrites of rat neocortical pyramidal neurons. *Journal of Physiology* (London), *505,* 605–616.
Segev, I., Rinzel, J., & Shepherd, G. M. (1995). *Theoretical foundation of dendritic function*. London, England: MIT Press.
Sjöström, P. J., & Häusser, M. (2006). A cooperative switch determines the sign of synaptic plasticity in distal dendrites of neocortical pyramidal neurons. *Neuron, 51,* 227–238.
Spruston, N. (2008). Pyramidal neurons: Dendritic structure and synaptic integration. *Nature Reviews Neuroscience, 9,* 206–221.
Spruston, N., & Kath, W. L. (2004). Dendritic arithmetic. *Nature Neuroscience, 7,* 567–569.
Spruston, N., Stuart, G., & Häusser, M. (2008). Chapter 14. Dendritic integration In *Dendrites*, Stuart, G., Spruston, N., & Häusser, M. (Eds.). Oxford: Oxford University Press.
Stuart, G. J., Dodt, H. U., & Sakmann, B. (1993). Patch-clamp recordings from the soma and dendrites of neurons in brain slices using infrared video microscopy. *Pflügers Archiv, 423,* 511–518.
Stuart, G. J., & Häusser, M. (1994). Initiation and spread of sodium action potentials in cerebellar Purkinje cells. *Neuron, 13,* 703–712.
Stuart, G. J., & Sakmann, B. (1994). Active propagation of somatic action potentials into neocortical pyramidal cell dendrites. *Nature, 367,* 69–72.
Stuart, G. J., Schiller, J., & Sakmann, B. (1997). Action potential initiation and propagation in rat neocortical pyramidal neurons. *Journal of Physiology* (London), *505,* 617–632.
Stuart, G. J., Spruston, N., Sakmann, B., & Häusser, M. (1997). Action potential initiation and backpropagation in neurons of the mammalian CNS. *Trends in Neurosciences, 20,* 125–131.
Takahashi, T. (1978). Intracellular recording from visually identified motoneurons in rat spinal cord slices. *Proceedings of the Royal Society of London Series B Biological Science, 202,* 417–421.
Tsubokawa, H., & Ross, W. N. (1996). IPSPs modulate spike backpropagation and associated [Ca^{2+}]i changes in the dendrites of hippocampal CA1 pyramidal neurons. *Journal of Neurophysiology, 76,* 2896–2906.
Waters, J., & Helmchen, F. (2004). Boosting of action potential backpropagation by neocortical network activity *in vivo*. *Journal of Neuroscience, 24,* 11127–11136.
Waters, J., Larkum, M., Sakmann, B., & Helmchen, F. (2003). Supralinear Ca2+ influx into dendritic tufts of layer 2/3 neocortical pyramidal neurons *in vitro* and *in vivo*. *Journal of Neuroscience, 23,* 8558–8567.
Williams, S. R., & Stuart, G. J. (2002). Dependence of EPSP efficacy on synapse location in neocortical pyramidal neurons. *Science, 295,* 1907–1910.

Suggested Readings

Häusser, M., & Mel, B. (2003). Dendrites: Bug or feature. *Current Opinion in Neurobiology, 13,* 372–383.
Häusser, M., Spruston, N., & Stuart, G. J. (2000). Diversity and dynamics of dendritic signaling. *Science, 290,* 739–744.
Hering, H., & Sheng, M. (2001). Dendritic spines: Structure, dynamics and regulation. *Nature Reviews Neuroscience, 2,* 880–888.
Hille, B. (2001). *Ion channels of excitable membranes*. Sunderland, MA: Sinauer.
Holtmaat, A., & Svoboda, K. (2009). Experience-dependent structural synaptic plasticity in the mammalian brain. *Nature Reviews Neuroscience, 10,* 647–658.
Jack, J. J. B. (1983). *Electric current flow in excitable cells*. (D. Noble, Ed.). Oxford: Clarendon Press.
Johnston, D. (1995). *Foundations of Cellular Neurophysiology*. (S. M.-S. Wu, Ed.). Cambridge, MA: MIT Press.
Johnston, D, & Narayanan, R. (2008). Active dendrites: colorful wings of the mysterious butterflies. *Trends in Neurosciences, 31,* 309–316.
Johnston, D., Magee, J. C., Colbert, C. M., Cristie, B. R. (1996). Active properties of neuronal dendrites. *Annu Rev Neurosci, 19,* 165–86.
Kampa, B. M., Letzkus, J. J., & Stuart, G. J. (2007). Dendritic mechanisms controlling spike-timing-dependent synaptic plasticity. *Trends in Neurosciences, 30,* 456–463.
Kasai, H., Matsuzaki, M., Noguchi, J., Yasumatsu, N., & Nakahara, H. (2003). Structure-stability-function relationships of dendritic spines. *Trends in Neurosciences, 26,* 360–368.
Lai, H. C., & Jan, L. Y. (2006). The distribution and targeting of neuronal voltage-gated ion channels. *Nature Reviews Neuroscience, 7,* 548–562.
London, M., & Häusser, M. (2005). Dendritic computation. *Annual Review of Neuroscience, 28,* 503–532.
Magee, J. C. (2000). Dendritic integration of excitatory synaptic input. *Nature Reviews Neuroscience, 1,* 181–190.
Nimchinsky, E. A., Sabatini, B. L., & Svoboda, K. (2002). Structure and function of dendritic spines. *Annual Review of Physiology, 64,* 313–353.
Schoppa, N. E., & Urban, N. N. (2003). Dendritic processing within olfactory bulb circuits. *Trends in Neurosciences, 26,* 501–506.
Segev, I., & Rall, W. (1998). Excitable dendrites and spines: Earlier theoretical insights elucidate recent direct observations. *Trends in Neurosciences, 21,* 453–460.
Segev, I., Rinzel, J., & Shepherd, G. M. (1995). *Theoretical foundation of dendritic function*. London, England: MIT Press.
Silver, R. A. (2010). Neuronal arithmetic. *Nature Reviews Neuroscience, 11,* 474–489.
Sjöström, P. J., Rancz, E. A., Roth, A., & Häusser, M. (2008). Dendritic excitability and synaptic plasticity. *Physiological reviews, 88,* 769–840.

Spruston, N. (2008). Pyramidal neurons: dendritic structure and synaptic integration. *Nature Reviews Neuroscience, 9,* 206–221.

Spruston, N., Jaffe, D. B., & Johnston, D. (1994). Dendritic attenuation of synaptic potentials and currents: The role of passive membrane properties. *Trends in Neurosciences, 17,* 161–166.

Stuart, G. J., Spruston, N., Sakmann, B., & Häusser, M. (1997). Action potential initiation and backpropagation in neurons of the mammalian CNS. *Trends in Neurosciences, 20,* 125–131.

Williams, S. R., & Stuart, G. J. (2003). Role of dendritic synapse location in the control of action potential output. *Trends in Neurosciences, 26,* 147–154.

Nelson Spruston, Michael Häusser, and Greg Stuart

CHAPTER

12

Brain Energy Metabolism

All the processes described in this textbook require energy. Ample clinical evidence indicates that the brain is exquisitely sensitive to perturbations of energy metabolism. This chapter covers the topics of energy delivery, production, and utilization by the brain. Careful consideration of the basic mechanisms of brain energy metabolism is an essential prerequisite to a full understanding of the physiology and pathophysiology of brain function. Abnormalities in brain energy metabolism are observed in a variety of pathological conditions such as, for example, neurodegenerative diseases, stroke, epilepsy, and migraine. The chapter reviews the features of brain energy metabolism at the global, regional, and cellular levels and extensively describes recent advances in the understanding of neuro-glial metabolic cooperation. A particular focus is the cellular and molecular mechanisms that tightly couple neuronal activity to energy consumption. This tight coupling is at the basis of functional brain-imaging techniques, such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI).

ENERGY METABOLISM OF THE BRAIN AS A WHOLE ORGAN

Glucose Is the Main Energy Substrate for the Brain

The human brain constitutes only 2% of the body weight, yet the energy-consuming processes that ensure proper brain function account for approximately 25% of total body glucose utilization. With a few exceptions that will be reviewed later, glucose is the obligatory energy substrate of the brain. In any tissue, glucose can follow various metabolic pathways; in the brain, glucose is almost entirely oxidized to CO_2 and water through its sequential processing by glycolysis (Fig. 12.1), the tricarboxylic acid (TCA) cycle also referred to as the Krebs cycle (Fig. 12.2), and the associated oxidative phosphorylation, which yield, on a molar basis, between 30 and 36 ATPs per glucose, depending on the coupling efficiency of oxidative phosphorylation. Indeed, the oxygen consumption of the brain, which accounts for almost 20% of the oxygen consumption of the whole organism, is 160 mmol per 100 g of brain weight per minute and roughly corresponds to the value determined for CO_2 production. This O_2/CO_2 relation corresponds to what is known in metabolic physiology as a respiratory quotient of nearly 1 and demonstrates that carbohydrates, and glucose in particular, are the exclusive substrates for oxidative metabolism.

This rather detailed information of whole brain energy metabolism was obtained using an experimental approach in which the concentration of a given substrate in the arterial blood entering the brain through the carotid artery is compared with that present in the venous blood draining the brain through the jugular vein. If the substrate is utilized by the brain, the arteriovenous (A-V) difference is positive; in certain cases, the A-V difference may be negative, indicating that metabolic pathways resulting in the production of the substrate predominate. In addition, when the rate of cerebral blood flow (CBF) is known, the steady-state rate of utilization of the substrate can be determined per unit time and normalized per unit brain weight according to the following relation: CMR = CBF (A-V), where CMR is the cerebral metabolic rate of a given substrate.

This approach was pioneered by Seymour Kety and C. F. Schmidt in the late 1940s and was further developed in the 1950s and 1960s. In normal adults, CBF is approximately 57 ml per 100 g of brain weight per minute, and the calculated glucose utilization by the brain is 31 mmol per 100 g of brain weight per minute, as determined with the A-V difference method. This value is slightly higher than that predicted from the rate of oxygen consumption

© 2013 Elsevier Inc. All rights reserved.

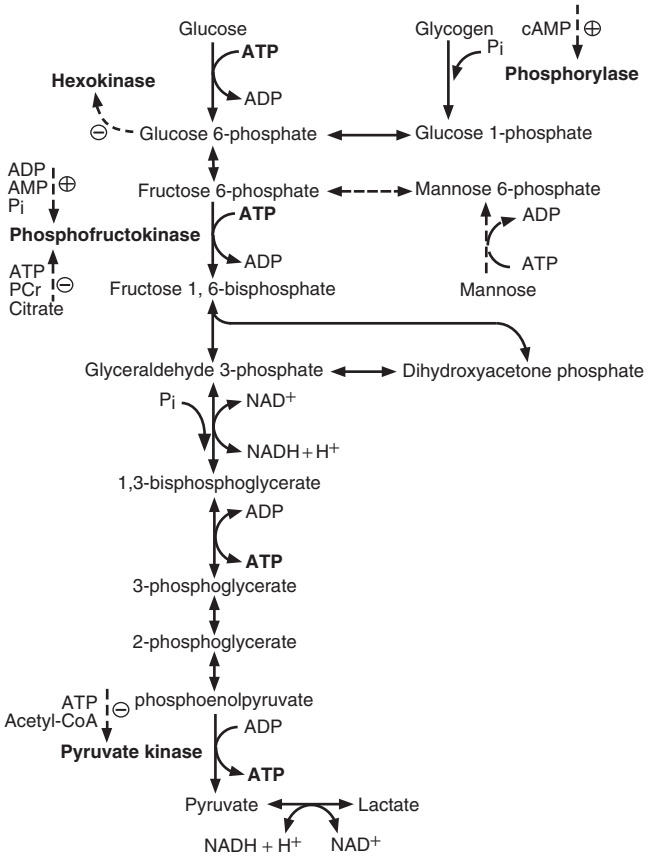


FIGURE 12.1 Glycolysis (Embden-Meyerhof pathway). Glucose phosphorylation is regulated by hexokinase, an enzyme inhibited by glucose 6-phosphate. Glucose must be phosphorylated to glucose 6-phosphate to enter glycolysis or to be stored as glycogen. Two other important steps in the regulation of glycolysis are catalyzed by phosphofructokinase and pyruvate kinase. Their activities are controlled by the levels of high-energy phosphates, as well as of citrate and acetyl-CoA. Pyruvate, through lactate dehydrogenase, is in dynamic equilibrium with lactate. This reaction is essential to regenerate NAD^+ residues necessary to sustain glycolysis downstream of glyceraldehyde 3-phosphate. PCr, phosphocreatine.

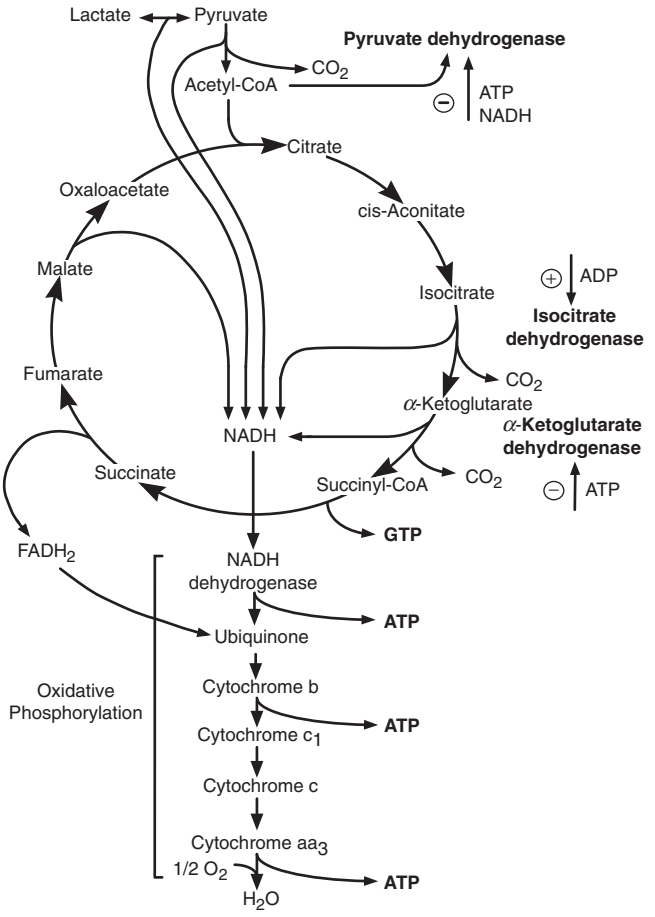


FIGURE 12.2 Tricarboxylic acid cycle (TCA cycle) and oxidative phosphorylation. Pyruvate entry into the cycle is controlled by pyruvate dehydrogenase activity that is inhibited by ATP and NADH. Two other important regulatory steps in the cycle are controlled by isocitrate and α-ketoglutarate dehydrogenases, whose activities are controlled by the levels of high-energy phosphates.

of the brain. Thus, in an organ such as the brain with a respiratory quotient of 1, the stoichiometry would predict that 6 mmol of oxygen are needed to fully oxidize 1 mmol of the six-carbon molecule of glucose; given an oxygen consumption rate of 160 mmol per 100 g of brain weight per minute, the predicted glucose utilization would be 26.6 mmol per 100 g of brain weight per minute (160:6), yet the actual measured rate is 31 mmol. What then is the fate of the excess 4.4 mmol? First, glucose metabolism may proceed, to a very limited extent, only through glycolysis, resulting in the production of lactate without oxygen consumption (Fig. 12.1); glucose can also be incorporated into glycogen (Fig. 12.1). Second, glucose is an essential constituent of macromolecules such as glycolipids and glycoproteins present in neural cells. Finally, glucose enters the metabolic pathways that result in the synthesis of three key neurotransmitters of the brain: glutamate, GABA, and acetylcholine (see Chapter 6).

Ketone Bodies Become Energy Substrates for the Brain under Particular Circumstances

Under particular circumstances, substrates other than glucose can be utilized by the brain. For example, breastfed neonates have the capacity to utilize the ketone bodies acetoacetate (AcAc) and D-3-hydroxybutyrate (3-HB) or free fatty acids, in addition to glucose, as energy substrates for the brain. This capacity is an interesting example of a developmentally regulated adaptive mechanism because maternal milk is highly enriched in lipids, resulting in a

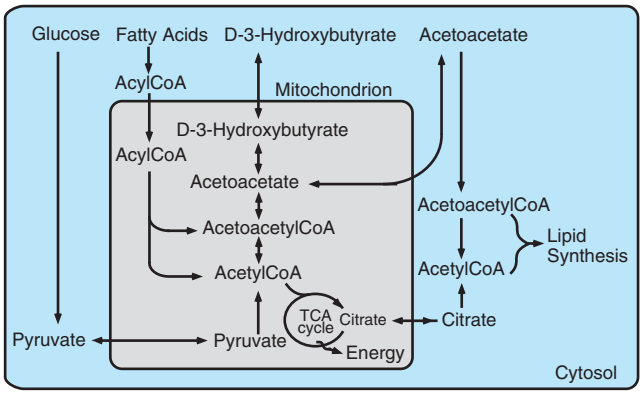


FIGURE 12.3 Relationship between lipid metabolism and the TCA cycle. Under particular dietary conditions, such as lactation in newborns or fasting in adults, the ketone bodies acetoacetate and D-3-hydroxybutyrate and circulating fatty acids can provide substrates to the TCA cycle after conversion to acetyl-CoA. Carbon atoms for lipid synthesis can be provided by glucose through citrate produced in the TCA cycle, a particularly relevant process for the developing brain.

lipid-to-carbohydrate ratio much higher than that present in postweaning nutrients. Indeed, lipids account for approximately 55% of the total calories contained in human milk, in contrast with 30–35% for a balanced postweaning diet. Acetoacetate, 3-HB, and free fatty acids can all be processed to acetyl-CoA, thus providing ATP through the TCA cycle (Fig. 12.3).

Another consideration regarding the lipid-rich diet provided during the suckling period relates to its contribution to the process of myelination. Evidence shows that brain lipids can be synthesized from blood-borne precursors such as ketone bodies. In addition, when suckling rats are fed a diet low in ketones, carbon atoms for lipogenesis can also be provided by glucose (Fig. 12.3). To summarize, ketone bodies are energy substrates, as well as precursors for lipogenesis during the suckling period; however, the developing brain appears to be metabolically quite flexible because glucose, in addition to its energetic function, can be metabolized to generate substrates for lipid synthesis.

Starvation and diabetes are two situations in which the availability of glucose to tissues is inadequate and in which plasma ketone bodies are elevated because of enhanced lipid catabolism. Under these conditions, the adaptive mechanisms described for breast-fed neonates become operative in the brain, allowing it to utilize AcAc or 3-HB as energy substrates.

Lactate and Pyruvate Serve as Instructive Cases

In addition to ketone bodies, lactate and pyruvate represent energy substrates alternative to glucose for neural cells. These two monocarboxylates can readily enter the cells through specialized monocarboxylate transporters (MCTs) and can be metabolized in the mitochondria through the TCA cycle and the oxidative phosphorylation to produce energy in the form of ATP (Fig. 12.2). *In vitro* and *ex vivo* data have provided extensive demonstration that lactate and pyruvate can sustain neuronal function—that is—synaptic activity, even in the absence of glucose (Schurr, 2006). In line with this, both monocarboxylates can protect neurons from cell damage/death in conditions of energy deprivation such as in hypoglycaemia.

It was initially thought that the permeability of the blood-brain barrier to monocarboxylates was limited, suggesting that circulating lactate or pyruvate could not represent significant alternative substrates to glucose for the brain *in vivo*. The demonstration of the expression of MCTs on intraparenchymal brain capillaries along with magnetic resonance spectroscopy (MRS) experiments have led to the reevaluation of the importance of circulating monocarboxylates, and in particular lactate, for brain energy metabolism. It is now estimated that plasma lactate (≈1.0 mM) is taken up by the human brain, where it is fully oxidized accounting for up to 8 to 10% of its energy requirements. This contribution may even be greater at supraphysiological lactate concentrations (i.e., following intravenous lactate infusion) with a near linear relationship. Increases in blood lactate concentration are observed in different physiological conditions such as during physical exercise, implying that lactate can readily be used by the brain as an alternative energy substrate in such specific conditions. In particular, during moderate-to-vigorous exercises, which results in blood lactate concentration up to 10 mM, the human brain takes up and oxidizes lactate to a considerably larger extent than under normal conditions (which could cover up to 20 to 25% of the total brain energy demand); this increase in lactate oxidation occurs at the expense of blood glucose utilization. Overall these data demonstrate that plasma lactate can be an energy substrate for the human brain. In addition, if formed within the brain parenchyma from glucose that has crossed the blood-brain barrier, lactate and pyruvate may in fact become the preferential energy substrates for activated neurons (see below).

Summary

Glucose is the obligatory energy substrate for brain, and it is almost entirely oxidized to CO_2 and H_2O. This simple statement summarizes, with few exceptions, over five decades of careful studies of brain energy metabolism at organ and regional levels. Under ketogenic conditions, such as starvation and diabetes and during breastfeeding, ketone bodies may provide an energy source for the brain. Lactate, formed from glucose within the brain parenchyma or imported from the circulation, is an adequate energy substrate as well.

TIGHT COUPLING OF NEURONAL ACTIVITY, BLOOD FLOW, AND ENERGY METABOLISM

A striking characteristic of the brain is its high degree of structural and functional specialization. Thus, when we move an arm, motor areas and their related pathways are activated selectively (see Chapter 27); intuitively, one can predict that as "brain work" increases locally (e.g., in motor areas), the energy requirements of the activated regions will increase in a temporally and spatially coordinated manner. Because energy substrates are provided through the circulation, blood flow should increase in the modality-specific activated area. In agreement with this prediction, a fascinating and robust feature of brain energy metabolism is the existence of such a tight coupling between energy demand and supply. Indeed, task-dependent increases in cerebral activity are invariably accompanied by changes in local cerebral blood flow and glucose utilization—these processes being referred to as neurovascular- and neurometabolic-coupling, respectively (Fig. 12.4A and Box 12.1).

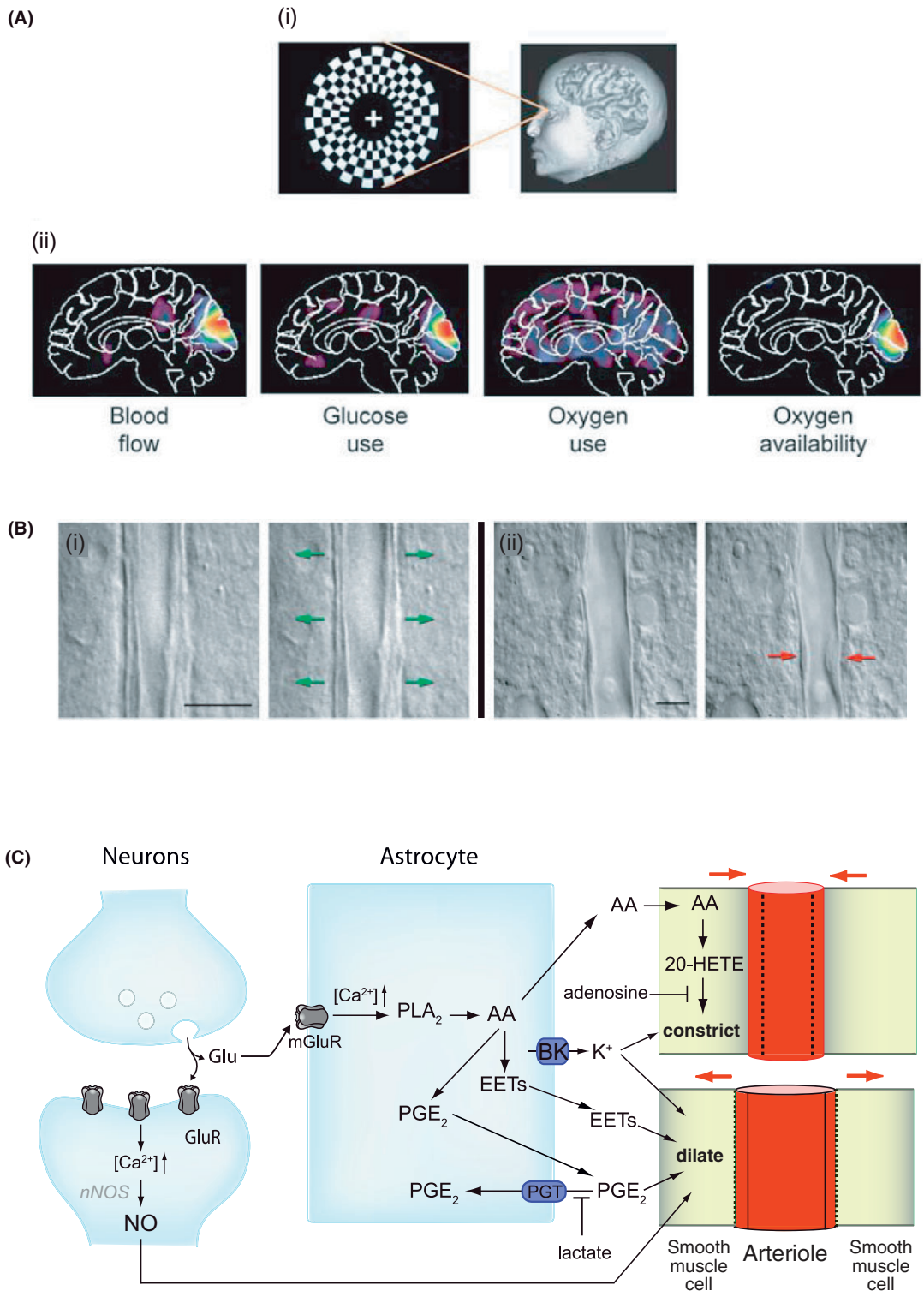
Which Mechanisms Couple Neuronal Activity to Blood Flow?

More than a century ago, the British neurophysiologist Charles Sherrington showed, in experimental animals, increases in blood flow localized to the parietal cortex in response to sensory stimulation. He postulated that "the brain possesses intrinsic mechanisms by which its vascular supply can be varied locally in correspondence with local variations of functional activity." With remarkable insight, he also proposed that "chemical products of cerebral metabolism" produced in the course of neuronal activation could provide the mechanism to couple activity with increased CBF (also known as functional hyperemia).

Since Sherrington's seminal work, a variety of vasoactive agents, including H^+, K^+, neurotransmitters (such as noradrenaline, serotonin, or vasoactive intestinal peptide [VIP]), adenosine, lactate, arachidonic acid metabolites, and nitric oxide (NO), released from different cellular compartments (i.e., neurons, astrocytes, pericytes, and smooth muscle cells) have been implicated in the activity-dependent increase in CBF. In conjunction with this, evidence suggests that the vascular tone is dynamically regulated, since both dilation and constriction of microvessels can be elicited in *ex vivo* preparations (Fig. 12.4B) (Attwell et al., 2010; Cauli et al., 2004; Petzold & Murthy, 2011). Although the exact physiological mechanisms underlying regulation of the vascular tone, and in particular functional hyperemia, are still incompletely understood, the importance played by neuron-to-astrocyte and glutamate signaling in this process is now widely demonstrated (Fig. 12.4C).

Nitric oxide has been recognized for a long time as an essential mediator for coupling neuronal activity to blood flow. NO is mainly formed locally by neurons under the action of glutamate (Fig. 12.4C), the principal excitatory neurotransmitter. Nitric oxide is a diffusible and potent vasodilator whose short half-life spatially and temporally restricts its domain of action. However, in several experimental models in which the activity of NO synthase, the enzyme responsible for NO synthesis, was inhibited, a certain degree of coupling was still observed, indicating that NO is only one of the regulators of local blood flow acting in synergy with others.

FIGURE 12.4 Neurovascular and neurometabolic coupling. (A) Differential regulation of blood flow (CBF), oxygen use ($CMRO_2$), and glucose use (CMRglu) during brain activation in humans. Stimulation of the human visual cortex was performed experimentally by presenting a visual stimulus in the form of a reversing annular checkerboard (i), and CBF, $CMRO_2$, and CMRglu were determined by PET (ii) (see Box 12.1). When compared with viewing a blank screen, visual stimulation produces marked changes in activity in visual areas of the brain, as shown in PET images using pseudo colors (from small increase, blue, to major increase, red). Increases in both blood flow and glucose use in the visual cortex could be observed that are unaccompanied by similar increases in oxygen use, suggesting that the metabolic needs associated with brain activation are met by glycolysis. As a result, the local availability of oxygen increases because the increased supply of oxygen by increased (arterial) blood flow exceeds the increased local demand for oxygen. Such an effect may account for the blood-oxygen-level-dependent (BOLD) signal of functional magnetic resonance imaging (fMRI). *Reproduced with permission from Raichle and Mintun (2006).* (B) Dynamic regulation of vascular tone. Single interneuron stimulation is sufficient to either dilate (i) or constrict (ii) neighboring microvessels in acute rat cortical slices. In (i) and (ii), images of blood vessels before (left) and after (right) electrical stimulation of the interneurons are shown. Scale bars, 10 μm. *Reproduced with permission from Cauli et al. (2004).* (C) Glutamate-dependent regulation of CBF. Neuronal activity-dependent glutamate release raises $[Ca^{2+}]_i$ in astrocytes by activating metabotropic glutamate receptors (mGluR). Ca^{2+} transients in astrocytes activate cytosolic phospholipase A_2 (PLA_2), thus producing arachidonic acid (AA). AA can generate different types of vasodilatory metabolites in astrocytes such as prostaglandins (PG), in particular PGE_2, and epoxyeicosatrienoic acids (EETs). In parallel, AA can diffuse to arteriolar smooth muscle cells, where it is converted to 20-hydroxyeicosatetraenoic acid (20-HETE) by ω-hydroxylase, which exerts a vasoconstricting effect. During brain activation, vasodilation is promoted by the increases in extracellular lactate and adenosine concentrations. Indeed, lactate favors PGE_2 accumulation in the extracellular space, and thus vasodilation, by decreasing the efficacy of the prostaglandin transporter (PGT) in astrocytes, whereas adenosine has an inhibitory effect on 20-HETE-mediated arteriolar constriction. In parallel to this, Ca^{2+} induces K^+ release from astrocytes through the activation of calcium-dependent potassium channels (BK), which in turn can promote either vasodilation or vasoconstriction, depending on the level of extracellular K^+. In postsynaptic neurons, glutamate activates ionotropic receptors, which trigger Ca^{2+} elevations, activation of the neuronal nitric oxide synthase (nNOS) and the release of the vasodilating nitric oxide (NO). Additional mechanisms are known to participate in CBF regulation but are not included in the present figure for clarity. A detailed description of all these mechanisms can be found elsewhere (see, for example, Attwell et al., 2010; Petzold & Murthy, 2011). *Modified with permission from Bélanger et al. (2011).*



More recent *in vitro* and *in vivo* experiments have provided evidence that astrocytes may play a key role in neurovascular coupling. Indeed, as will be elaborated in greater detail in the section(s) devoted to neurometabolic coupling (e.g., see Fig. 12.11), astrocytes occupy a strategic position between capillaries and the neuropil. Through receptors and reuptake sites for neurotransmitters, notably glutamate, they can sense synaptic activity and couple it to vascular response. In particular, calcium transients in astrocytes resulting from metabotropic

BOX 12.1

POSITRON EMISSION TOMOGRAPHY (PET) AND FUNCTIONAL MAGNETIC RESONANCE IMAGING (fMRI)

We have seen that neuronal activity is tightly coupled to blood flow and metabolism. With the advent of sophisticated imaging procedures such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI), it is now possible to detect the signals generated by the metabolic processes associated with neuronal activity, thus providing a unique opportunity to see the "brain at work." Indeed, local changes in blood flow, glucose utilization, and oxygen consumption can be monitored noninvasively under basal and activated conditions in human subjects. How is this possible?

For PET, a solution containing slightly radioactive molecules is injected into the circulation, and its sites of brain uptake can be visualized. The molecule is labeled radioactively with an unstable radionuclide possessing an excess number of protons; as a consequence of normal radioactive decay, the excess proton is converted to a neutron. In this process, a positron (a positively charged electron) is emitted and collides with an electron, releasing energy in the form of two photons with opposite trajectories. The photons are sensed by specialized detectors placed around the head; when two photons simultaneously reach two detectors positioned at 180 degrees of each other, the origin of the positron–electron collision can be localized with a millimeter resolution. Commonly used positron-emitting radionuclides are oxygen-15 (^{15}O), carbon-11 (^{11}C), and fluorine-18 (^{18}F). Blood flow is monitored with ^{15}O-labeled water and glucose utilization with ^{18}F-2-deoxyglucose. Oxygen consumption is visualized directly with ^{15}O. With the use of sophisticated algorithms to process data, the localization of activity to specific brain areas during a given task (sensory, motor, cognitive) can be achieved. Thus, the activity of neuronal ensembles, and of the associated glia, coupled to increased blood flow and glucose utilization results in a localized signal due to the augmented concentration of ^{15}O-labeled water (monitoring blood flow) and ^{18}F-2-deoxyglucose (assessing glucose utilization) in the activated area (see Fig. 12.4A).

PET is one of the ways in which brain work can be visualized. A now popular technique, fMRI relies on the magnetic signals detected in an activated brain region in relation to its degree of oxygenation. Depending on the degree to which it is saturated by oxygen, hemoglobin (by acting as a paramagnetic contrast agent) can alter the magnetic signal detected in a tissue exposed to the magnetic fields used for structural MRI. In other words, different MRI signals can be obtained depending on the oxyhemoglobin/deoxyhemoglobin ratio in a given brain area. Local activation of a brain area results in increased blood flow (see Fig. 12.4A). Although the precise mechanisms for the fMRI signals are still being discussed, it is currently thought that this phenomenon, by leading to a localized enrichment in oxyhemoglobin, alters the oxyhemoglobin/deoxyhemoglobin ratio, providing the blood-oxygen-level-dependent (BOLD) signal for fMRI. Functional MRI is a remarkably convenient and powerful technique: the signal acquisition time is extremely rapid (in the order of seconds), and the resolution equals that of PET (i.e., a few millimeters). In addition, fMRI is totally noninvasive and can thus be repeated frequently on the same subject, who can then serve as its own control. The model of neurometabolic coupling provided by the ANLS can account for the unexpected increases in oxyhemoglobin during activation, which is at the basis of the BOLD signal. Indeed, the Astrocyte-Neuron Lactate Shuttle (ANLS) predicts that a *transient nonoxidative* use of glucose occurs during the initial phases (1 to 2 seconds) of activation, resulting in lactate production by astrocytes. Since activity-dependent increase in CBF (neurovascular coupling) brings to the activated area oxygen-enriched (arterial) blood, while oxygen consumption does not increase commensurately because of the transient nonoxidative use of glucose in astrocytes (see Fig. 12.4A), it follows that the activated region will contain a higher ratio of oxy/deoxy hemoglobin at the basis of the BOLD signal. Recoupling with oxidative processing of lactate/pyruvate in neurons reestablishes the oxy/deoxyhemoglobin ratio, consistent with the transient nature of the kinetics for the BOLD signal.

glutamate receptor activation have been shown to induce the production and release of vasoactive agents from astrocytes which are associated with both dilation or constriction of blood vessels (Fig. 12.4C). These vasoactive agents include products of arachidonic acid (AA) metabolism such as (1) the two vasodilators epoxyeicosatrienoic acids (EETs) and prostaglandin E2 (PGE_2) formed within astrocytes, and (2) the vasoconstrictor 20-hydroxyeicosatetraenoic acid (20-HETE) that is produced in smooth muscle cells from AA released by astrocytes (Fig. 12.4C) (Attwell et al., 2010; Petzold & Murthy, 2011).

Interestingly, Gordon and collaborators demonstrated in rat brain slices that the balance between vasoconstriction

and vasodilation may be dictated by brain metabolism. They observed that neuronal activation leads to arteriolar dilatation under low-oxygen conditions, whereas in high-oxygen conditions the same stimulus induces the constriction of arterioles (see Attwell et al., 2010; Petzold & Murthy, 2011). The authors also demonstrated that the vasodilating effect observed in lower-oxygen conditions is mediated by the activation of glycolysis and an elevation of extracellular lactate and adenosine concentrations, which interfere with PGE_2 and 20-HETE metabolisms, respectively (Fig. 12.4C). This set of observations suggests that under resting conditions, when oxygen is not being rapidly consumed, astrocytic Ca^{2+} signals would induce a constricting tone, keeping cerebral blood flow at an appropriately lower level. In contrast, during activation, the drop in pO_2 due to oxygen consumption (see Attwell et al., 2010) and the rise in extracellular lactate and adenosine levels would promote astrocyte-mediated vasodilation. Another important vasoactive agent released by astrocytes, which is also coupled to elevation of calcium levels, is K^+. Indeed, activation of calcium-sensitive potassium (BK) channels leads to K^+ release into the extracellular space which triggers vasodilation, through hyperpolarization of smooth muscle cells (see Petzold & Murthy, 2011) (Fig. 12.4C). Interestingly, overactivation of BK channels produces supraphysiologic elevation of external $[K^+]$ (similar to those found in pathological conditions such as ischemia or spreading depression), which shifts vasodilation to vasoconstriction, suggesting that the ambient level of extracellular K^+ dictates the vascular tone.

In summary, several products of activity-dependent neuronal and glial metabolism such as NO, lactate, adenosine, prostanoids, and K^+ exert vasoactive effects and are thought to be major (but not exclusive when taken individually) mediators of hyperemia. Importantly, these mediators appear to act in concert to dynamically regulate, both spatially and temporally, the vascular tone to adapt energy demand and supply. Nitric oxide is undoubtedly a key element in coupling, particularly in view of the fact that glutamate, the principal excitatory neurotransmitter, triggers a receptor-mediated NO formation in neurons. In conjunction to this, astrocytes have emerged as essential intermediary processors in neurovascular coupling. Nevetheless, the complex mechanisms involved in the coupling of neuronal activation and increased CBF are not yet fully elucidated and still await firm functional confirmation *in vivo*.

Through the activity-linked increase in blood flow, more substrates—namely, glucose and oxygen—necessary to meet the additional energy demands are delivered to the activated area per unit time. The cellular and molecular mechanisms involved in oxygen consumption and glucose utilization are treated in later sections.

Blood Flow and Energy Metabolism Can Be Visualized in Humans

Modern functional brain-imaging techniques enable the *in vivo* monitoring of human blood flow and the two indices of energy metabolism: glucose utilization and oxygen consumption (Fig. 12.4A and Box 12.1) (Raichle & Mintun, 2006). For instance, with the use of PET and appropriate positron-emitting isotopes such as ^{18}F and ^{15}O, basal rates, as well as activity-related changes in local blood flow or oxygen consumption, can be studied using ^{15}O-labeled water or $^{15}O_2$, respectively (Fig. 12.4A). Local rates of glucose utilization [also defined as local cerebral metabolic rates for glucose (LCMRglu)] can be determined with ^{18}F-labeled 2-deoxyglucose (2-DG) (Fig. 12.4A). The use of 2-DG as a marker of LCMRglu was pioneered by Louis Sokoloff and associates at the National Institutes of Health, first in laboratory animals. The method is based on the fact that 2-DG crosses the blood-brain barrier, is taken up by brain cells, and is phosphorylated by hexokinase with kinetics similar to that for glucose; however, unlike glucose 6-phosphate, 2-deoxyglucose 6-phosphate cannot be metabolized further and therefore accumulates intracellularly (Fig. 12.5).

For studies in laboratory animals, tracer amounts of radioactive 2-DG are injected intravenously; the animal is subjected to the behavioral paradigms of interest and sacrificed at the end of the experiment. Serial thin sections of the brain are prepared and processed for autoradiography. This autoradiographic method provides, after appropriate corrections, an accurate measurement of LCMRglu with a spatial resolution of a few millimeters. Using this method, researchers have determined LCMRglu in virtually all structurally and functionally defined brain structures in various physiological and pathological states, including sleep, seizures, and dehydration, and after a variety of pharmacological treatments. Furthermore, glucose utilization increases in the pertinent brain areas during motor tasks or activation of pathways subserving specific modalities, such as visual, auditory, olfactory, or somatosensory stimulation. For example, in mice, sustained stimulation of the whiskers results in marked increases in LCMRglu in discrete areas of the primary sensory cortex called the barrel fields, where each whisker is represented with an extreme degree of topographical specificity (see Chapter 24). Basal glucose utilization of the gray matter as determined by 2-DG autoradiography varies, depending on the brain structure, between 50 and 150 mmol per 100 g of wet weight per minute in the rat.

In humans, LCMRglu determined by PET with the use of ^{18}F-2-DG is approximately 50% lower than that in rodents, and physiological activation of specific modalities increases LCMRglu in discrete areas of the brain that can be visualized with a spatial resolution of a few millimeters.

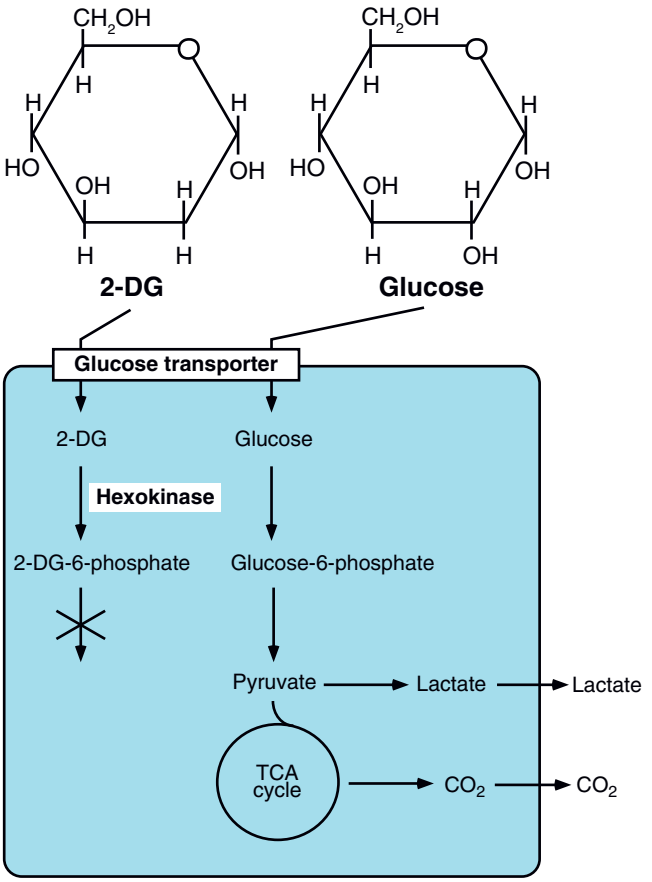


FIGURE 12.5 Structure and metabolism of glucose and 2-deoxyglucose (2-DG). 2-DG is transported into cells through glucose transporters and phosphorylated by hexokinase to 2-DG-6-phosphate without significant further processing or dephosphorylation back to glucose. Therefore, when labeled radioactively, 2-DG used in tracer concentrations is a valuable marker of glucose uptake and phosphorylation, which directly indicates glucose utilization.

For example, visual stimulations presented to subjects as checkerboard patterns reversing at frequencies ranging from 2 to 10 Hz selectively increase LCMRglu in the primary visual cortex and a few connected cortical areas. With the use of this stimulation paradigm, the combined PET analysis of local cerebral blood flow (LCBF) and local oxygen consumption ($LCMRO_2$), in addition to LCMRglu, has revealed a unique and unexpected feature of human brain energy metabolism regulation (Fig. 12.4A). The canonical view was that the three metabolic parameters were tightly coupled, implying that, if, for example, CBF increased locally during physiological activation, LCMRglu and $LCMRO_2$ would increase in parallel. In what is now referred to as the phenomenon of "uncoupling," physiological stimulation of the visual system increases LCBF and LCMRglu (both by 30 to 40%) in the primary visual cortex without a commensurate increase in $LCMRO_2$ (which increases only ≈5%) (Raichle & Mintun, 2006) (Fig. 12.4A), indicating that the additional glucose utilized during neuronal activation can be processed through glycolysis rather than through the TCA cycle and oxidative phosphorylation. The phenomenon of uncoupling has been confirmed in other cortical areas, although its magnitude may differ depending on the modality, and may actually be absent in certain cases. A glance at the metabolic pathways reveals that if glucose does not enter the TCA cycle to be oxidized, then lactate will be produced (Figs. 12.1 and 12.2).

Lactate, like several other metabolically relevant molecules, can be determined with the technique of magnetic resonance imaging (MRI) spectroscopy for ^1H, which provides a means of unequivocally identifying in living tissues the presence of molecules that bear the naturally occurring isotope ^1H. Consistent with the prediction that if during activation glucose is predominantly processed glycolytically, then lactate should be produced locally in the activated region, an activity-dependent transient increase in the lactate signal could be detected with ^1H MRI spectroscopy in different brain areas (Figley & Stroman, 2011, and references therein). Interestingly, recent evidence obtained in human visual cortex demonstrates that while ATP production positively correlates with $LCMRO_2$, task-induced CBF increase positively correlates with lactate production but negatively correlates with $LCMRO_2$ (see Figley & Stroman, 2011). This suggests that while energy demand is met through oxidative metabolism, CBF response is mediated by factors other than oxygen demand (see section above). These observations support the view that to face the local increases in energy demands linked to neuronal activation, the brain transiently resorts to an integrated sequence of glycolysis and oxidative phosphorylation (Magistretti & Pellerin, 1999). This transient uncoupling may vary in amplitude depending on the modalities of activation (Frackowiak et al., 2001) and is likely to occur in different cellular compartments—that is, astrocytes versus neurons (Figley & Stroman, 2011; Pellerin & Magistretti, 1994).

Summary

Studies at the whole organ level, based on the A-V differences of metabolic substrates, have revealed a great deal about the global energy metabolism of the brain. They have indicated that, under normal conditions, glucose is virtually the sole energy substrate for the brain and that it is entirely oxidized. New techniques that allow imaging of the three fundamental parameters of brain energy metabolism—namely, blood flow, oxygen consumption, and glucose utilization—provide a more refined level of spatial resolution and demonstrate that brain energy metabolism is regionally heterogeneous and is coupled tightly to the functional activation of specific neuronal pathways.

ENERGY-PRODUCING AND ENERGY-CONSUMING PROCESSES IN THE BRAIN

What are the cellular and molecular mechanisms that underlie the regulation of brain energy metabolism revealed by the foregoing studies at global and regional levels? In particular, what are the metabolic events taking place in the cell types that make up the brain parenchyma? How is it possible to reconcile whole organ studies indicating complete oxidation of glucose with transient activation-induced glycolysis at the regional level? These and other related questions will be addressed here and in the next sections.

Glucose Metabolism Produces Energy in the Form of ATP and Reducing Equivalents

Before we move on to an analysis of the cell-specific mechanisms of brain energy metabolism, it seems appropriate to briefly review some basic aspects of the energy balance of the brain. Because glucose, in normal circumstances, is the main energy substrate of the brain, the overview will be restricted to its metabolic pathways. Glucose metabolism in the brain is similar to that in other tissues and includes three principal metabolic pathways: (1) glycolysis occurring in the cytosol and associated with the pentose phosphate pathway and glucose storage in the form of glycogen (the latter only in astrocytes, see below), and (2) the TCA cycle and oxidative phosphorylation, which take place to the mitochondria. Because of the global similarities with other tissues, these pathways are simply summarized in Figures 12.1, 12.2, and 12.6, and only a few aspects specific to the nervous tissue will be discussed.

Glycolysis

Glycolysis (Embden-Meyerhof pathway) is the metabolism of glucose to pyruvate (Fig. 12.1). It results in the net production of only two molecules of ATP per glucose molecule; indeed, four ATPs are formed in the processing of glucose to pyruvate, whereas two ATPs are consumed to phosphorylate glucose to glucose 6-phosphate and fructose 6-phosphate to fructose 1,6-bisphosphate, respectively (Fig. 12.1). Under anaerobic conditions (meaning in the absence of oxygen), pyruvate is converted to lactate, allowing the regeneration of nicotinamide adenine dinucleotide (NAD^+), which is essential to maintain a continued glycolytic flux. Indeed, if NAD^+ were not regenerated, glycolysis could not proceed beyond glyceraldehyde 3-phosphate (Fig. 12.1). Another situation in which the end-product of glycolysis is lactate rather than pyruvate is when oxygen consumption does not match glucose utilization, implying that the rate of pyruvate production through glycolysis exceeds pyruvate oxidation by the TCA cycle (Fig. 12.2). This condition has been well described in skeletal muscle during intense exercise and in tumoral tissues, where it is referred to as the "Warburg effect". The occurrence of aerobic glycolysis in the nervous system is consistent with the transient uncoupling observed between glucose utilization and oxygen consumption that has been described in the human cerebral cortex during activation with the use of PET (Raichle & Mintun, 2006).

Tricarboxylic Acid Cycle

Under aerobic conditions, pyruvate is oxidatively decarboxylated to yield acetyl-CoA in a reaction catalyzed by the enzyme pyruvate dehydrogenase (PDH). Acetyl-coenzyme A condenses with oxaloacetate to produce citrate (Fig. 12.2). This is the first step of the TCA cycle, in which three pairs of electrons are transferred from NAD^+ to NADH—and one pair from flavin adenine dinucleotide (FAD) to its reduced form ($FADH_2$)—through four oxidation-reduction steps (Fig. 12.2). During the process of pyruvate oxidation, carbon atoms leave the TCA cycle in the form of CO_2. NADH and $FADH_2$ transfer their electrons to molecular O_2 through the mitochondrial electron transfer chain to produce 28-34 ATPs and H_2O in the process of oxidative phosphorylation. Thus, under aerobic conditions (i.e., when glucose is fully oxidized through the TCA cycle to CO_2 and H_2O), NAD^+ is regenerated enabling the TCA cycle to proceed further to generate ATP.

Pentose Phosphate Pathway

Although glycolysis, the TCA cycle, and oxidative phosphorylation are coordinated pathways that produce ATP, using glucose as a fuel, ATP is not the only form of metabolic energy. Indeed, for several biosynthetic reactions (anabolic reactions) in which the precursors are in a more oxidized state than the products, metabolic energy in the form of reducing power is needed in addition to ATP. This is the case for the reductive synthesis of free fatty acids from acetyl-CoA, which are components of myelin and of other structural elements of neural cells, such as the plasma membrane. In brain cells, as in other organs, the reducing power for anabolic reactions is provided by NADPH (the reduced form of nicotinamide adenine dinucleotide phosphate, $NADP^+$) which is mainly produced by glucose processing through the pentose phosphate pathway (PPP) (Fig. 12.6). The NADP/NADPH ratio is the single most important factor regulating the entry of glucose 6-phosphate into the PPP. Thus, when high reducing power is mobilized (see also below), NADPH levels decrease, and the pentose phosphate

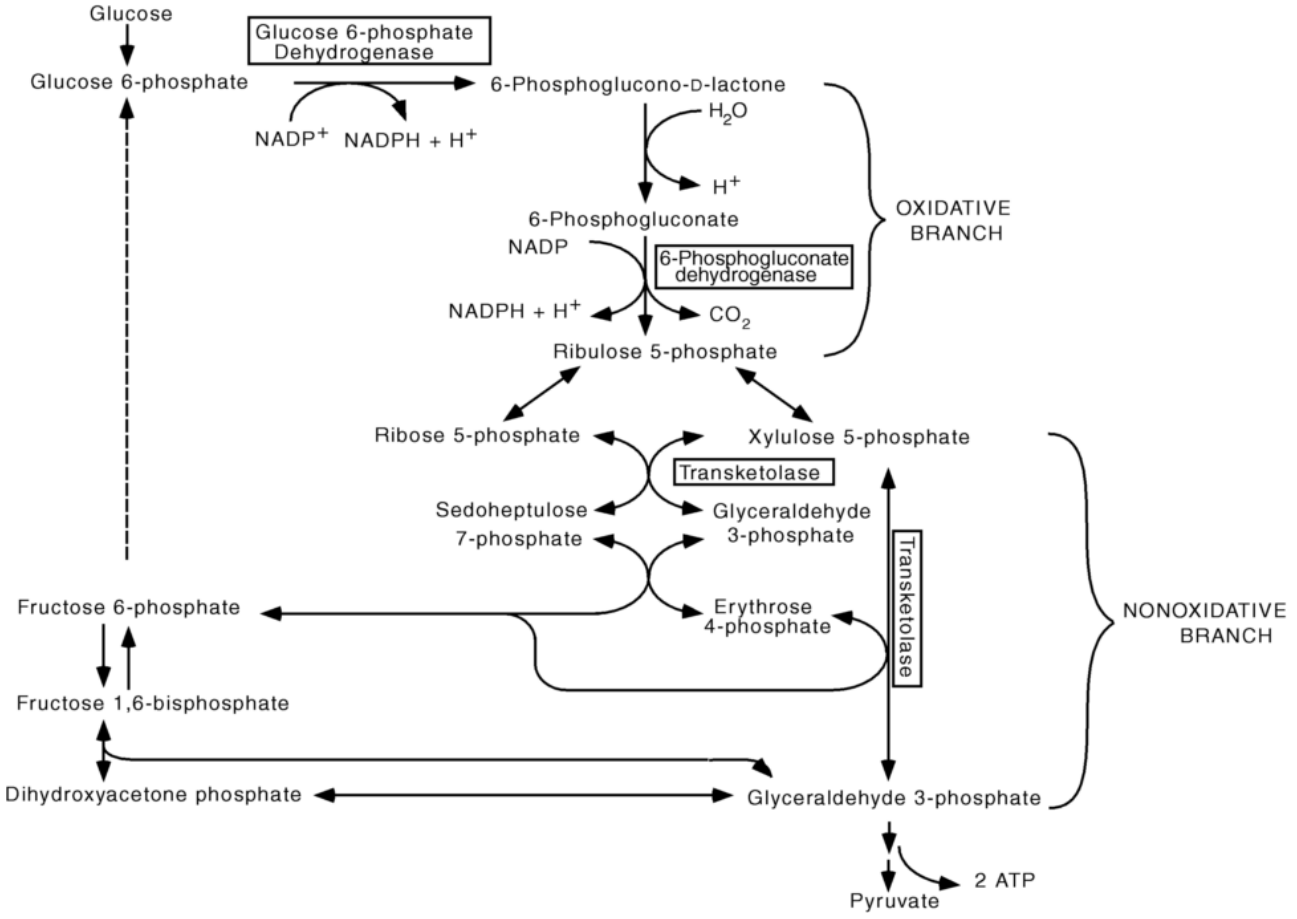


FIGURE 12.6 The pentose phosphate pathway. In the oxidative branch of the pentose phosphate pathway, two NADPH are generated per glucose 6-phosphate. The first rate-limiting reaction of the pathway is catalyzed by glucose 6-phosphate dehydrogenase; the second NADPH is generated through the oxidative decarboxylation of 6-phosphogluconate, a reaction catalyzed by 6-phosphogluconate dehydrogenase. The nonoxidative branch of the pentose phosphate pathway provides a reversible link with glycolysis by regenerating the two glycolytic intermediates glyceraldehyde 3-phosphate and fructose 6-phosphate. This regeneration is achieved through three sequential reactions. In the first, catalyzed by transketolase, xylulose 5-phosphate, and ribose 5-phosphate (which originate from ribulose 5-phosphate, the end product of the oxidative branch) yield glyceraldehyde 3-phosphate and sedoheptulose 7-phosphate. Under the action of transaldolase, these two intermediates yield fructose 6-phosphate and erythrose 4-phosphate. This latter intermediate combines with xylulose 5-phosphate, in a reaction catalyzed by transketolase, to yield fructose 6-phosphate and glyceraldehyde 3-phosphate. Thus, through the nonoxidative branch of the pentose phosphate pathway, two hexoses (fructose 6-phosphate) and one triose (glyceraldehyde 3-phosphate) of the glycolytic pathway are regenerated from three pentoses (ribulose 5-phosphate).

pathway is activated to generate new reducing equivalents. The pentose phosphate pathway is also tightly connected to glycolysis through two enzymes—transketolase and transaldolase—which recycle PPP metabolites to fructose 6-phosphate and glyceraldehyde 3-phosphate, two intermediates of glycolysis (Fig. 12.6).

Glucose Metabolism, Reactive Oxygen Species, and the Protective Role of Glutathione

In addition to anabolic reactions, NADPH is needed for the scavenging of reactive oxygen species (ROS). The superoxide radical anion (O_2^{-}.), hydrogen peroxide (H_2O_2), and the hydroxy radical (HO.) are three ROS, generated by the transfer of single electrons to molecular oxygen as by-products of several physiological cellular processes. A considerable contribution to the generation of ROS is the oxidative metabolism of glucose taking place in the mitochondrial electron transfer chain associated with oxidative phosphorylation. Other ROS-generating reactions include the activities of monoamine oxidase, tyrosine hydroxylase, nitric oxide synthase, and the eicosanoid-forming enzymes lipoxygenases and cyclooxygenases. Reactive oxygen species are highly damaging to cells because they can cause DNA disruption and mutations, as well as activation of enzymatic cascades, including proteases and lipases that can eventually lead to cell death.

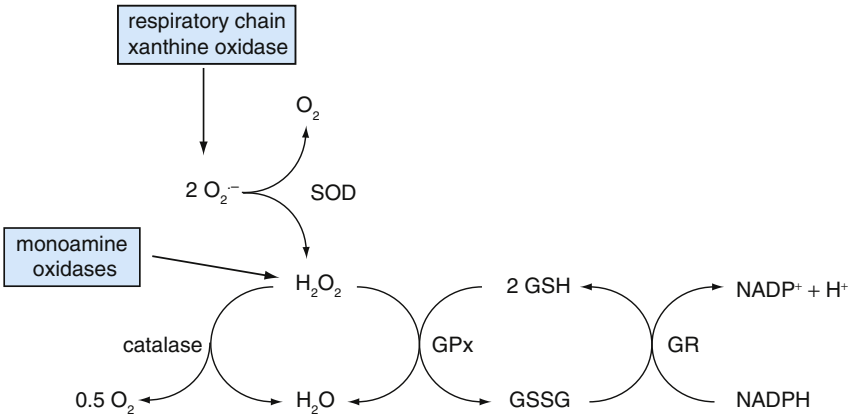


FIGURE 12.7 Enzymatic reactions for scavenging reactive oxygen species (ROS). The toxic superoxide anion ($O_2^{\cdot-}$) formed by a variety of physiological reactions, including oxidative phosphorylation (i.e., respiratory chain), is scavenged by superoxide dismutase (SOD), which converts the superoxide anion to hydrogen peroxide (H_2O_2) and molecular oxygen. Glutathione peroxidase (GPx) converts the still toxic hydrogen peroxide to water; reduced glutathione (GSH) is required for this reaction, in which it is converted to its oxidized form (GSSG). GSH is regenerated through the action of glutathione reductase (GR), a reaction requiring NADPH.

The coordinated activity of two molecules is essential in protecting cells against ROS-mediated damage, or oxidative stress: NADPH and glutathione. As we have seen, NADPH is produced through a particular arm of glucose metabolism, the PPP. Interestingly, therefore, glucose metabolism provides two forms of energy: high energy phosphates such as ATP and reducing power such as NAD(P)H, the latter contributing to the neutralization of ROS, the harmful by-products of the process (oxidative phosphorylation) that produces the former. It is, however, through its combined action with glutathione that NADPH contributes to ROS scavenging. Scavenging of ROS is ensured by the sequential action of superoxide dismutase (SOD) and glutathione peroxidase (Fig. 12.7). Thus, two superoxide anions formed by the aforementioned cellular processes are converted by SOD into H_2O_2, still a ROS. Glutathione peroxidase converts H_2O_2 into H_2O and O_2 at the expense of reduced glutathione (GSH), which is regenerated by glutathione reductase in the presence of NADPH (Fig. 12.7).

Glutathione is a tripeptide (GSH; γ-L-glutamyl-L-cysteinylglycine) synthesized through the concerted action of two enzymes: γGluCys synthase, which combines glutamate and cysteine to yield the dipeptide γGluCys, and glutathione synthase, which adds a glycine to the dipeptide to yield GSH (Fig. 12.8). The glutathione content and reducing potential are considerably higher in astrocytes compared to neurons; this fact, combined with the much higher oxidative activity of neurons versus astrocytes, makes neurons more vulnerable to oxidative stress as well as highly dependent on astrocytes for their protection. Indeed, among the numerous metabolic cooperations that exist between astrocytes and neurons (which will be developed in more detail in the next section), an important one with regard to oxidative stress defense mechanisms involves glutathione metabolism through the shuttling of GSH from astrocytes to neurons. Astrocytes release large amounts of GSH into the extracellular space; GSH is cleaved by the sequential action of ectoenzymes to Cys and Gly (Fig. 12.8). Cys and Gly are transported into neurons (note that neurons cannot take up GSH), providing two precursors for GSH synthesis. Glutamate, the third precursor of GSH, is also provided by astrocytes to neurons under the form of glutamine, from which glutamate is produced through the action of glutaminase, through the so-called glutamate-glutamine shuttle which will be described in the next section (Figs. 12.8 and 12.10).

Processes Linked to Neuronal Function Consume Energy

The main energy-consuming process of the brain is the maintenance of ionic gradients across the plasma membrane, a condition that is crucial for excitability. Indeed, neuronal signaling is based on the rapid flow of charges—carried by ions such as sodium, potassium, and calcium—across the membranes of the axon, dendrites, and soma. This is not expected to cost much *per se*, as the flow of ions follows favorable electrochemical gradients; however, reestablishing these constantly dissipated gradients, through the activity of ATP-consuming pumps such as the Na^+/K^+-ATPase, does indeed cost energy. As extensively reviewed in Chapters 5 and 10, two principal processes operate the signaling between neurons: the action potentials (AP) mainly carrying signals along the axon from the soma to the synapse (some APs, defined as backpropagating, are also generated in dendrites) and postsynaptic potentials (PSP), generated by synaptically released neurotransmitters acting at

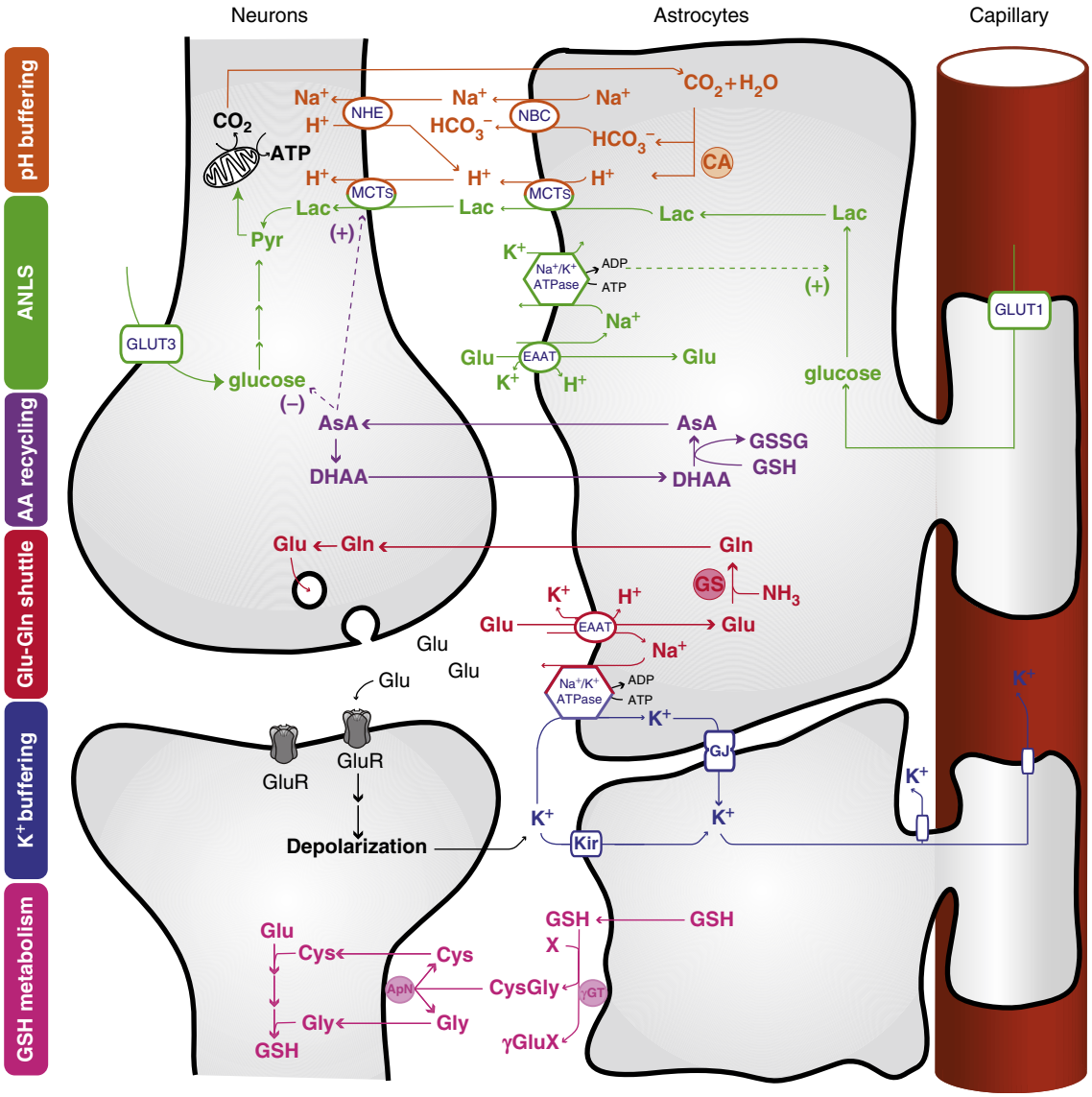


FIGURE 12.8 Central role of astrocytes in brain homeostasis. *Orange: pH buffering*. Abundant carbonic anhydrase (CA) in astrocytes converts CO_2 to H^+ and HCO_3^-. Two HCO_3^- are transported into the extracellular space along with one Na^+ via the Na^+-HCO_3^- cotransporter (NBC), thereby increasing the extracellular buffering power. Protons left in the glial compartment may drive the transport of lactate (Lac) outside of astrocytes and into neurons through monocarboxylate transporters (MCTs). Excess H^+ in neurons is extruded via sodium-hydrogen exchange (NHE). *Green: Astrocyte-neuron lactate shuttle (ANLS)*. Glutamate (Glu) uptake by astrocytes is accompanied by Na^+ entry, which is extruded by the action of the Na^+/K^+ ATPase. This triggers anaerobic glucose utilization in astrocytes and glucose uptake from the circulation through the glucose transporter GLUT1. The lactate produced is shuttled to neurons through MCTs, where it can be used as an energy substrate after its conversion to pyruvate (Pyr). Neurons can also take up glucose via the neuronal glucose transporter 3 (GLUT3). *Purple: Ascorbic acid recycling*. Astrocytes recycle oxidized ascorbic acid (dehydroascorbic acid; DHAA) back to ascorbic acid (AsA) using reduced glutathione (GSH) and forming oxidized glutathione (GSSG). AsA can be used directly in astrocytes or released and taken up by neurons for reactive oxygen species (ROS) scavenging. AsA may also act as a metabolic switch by inhibiting glucose consumption and stimulating lactate uptake in neurons. *Red: Glutamate-glutamine cycle.* Glutamate released into the synaptic cleft activates ionotropic glutamatergic receptor (GluR), producing a postsynaptic depolarization. Astrocytic excitatory amino acid transporters (EAATs) are responsible for the uptake of a large fraction of glutamate at the synapse. Glutamate is converted into glutamine (Gln) by glutamine synthase (GS) and shuttled back to neurons for glutamate resynthesis. *Blue: K^+ buffering*. Astrocytes buffer excess K^+ released into the extracellular space as a result of neuronal activity (for example, through inwardly rectifying K^+ channels, Kir). K^+ diffuses through the astrocytic network via gap junctions (GJ) down its concentration gradient and is released in sites of lower concentration. *Pink: Glutathione metabolism.* Astrocytes release glutathione (GSH) in the extracellular space, where it is cleaved by the astrocytic ectoenzyme γ-glutamyl transpeptidase (γGT). The resulting dipeptide CysGly is cleaved by the neuronal ectopeptidase aminopeptidase N (ApN), forming Cysteine (Cys) and Glycine (Gly), which serve as precursors for neuronal GSH synthesis. X represents an acceptor for the γ-glutamyl moiety (γGlu) in the reaction catalyzed by γGT. A detailed description of these neuron-astrocyte metabolic cooperations can be found elsewhere (see, for example, Bélanger & Magistretti, 2009, and references therein). *Reproduced with permission from Allaman et al. (2011).*

specific receptors present on dendrites and soma. Which of the AP and the PSP is more energetically costly? This question has been at the center of a long-standing debate in neuroscience. Early estimates in 1980s suggested that the energetic cost of signaling in the brain was dominated by PSPs. More recently, bottom-up energy budgets, in which the costs of each step of signaling were estimated, have proposed that APs and PSPs equally contribute to the energy consumption associated with neuronal activity in rodents (Attwell & Iadecola, 2002).

This is not just a question that excites neurophysiology *aficionados*: it has direct relevance for the interpretation of functional brain imaging signals because of the tight coupling between neuronal signaling and energy utilization. Experiments in which neuronal activity has been recorded while acquiring imaging data suggest that local field potentials, which reflect PSPs, best correlate with the imaging signal. In a series of elegant studies carried out in the rat hippocampus, Alle and colleagues have shown that the relative kinetic properties of sodium and potassium channels present on axons optimize the energy utilization necessary for APs generation. As described in Chapter 5, APs are produced by the sequential opening of voltage-sensitive sodium channels, resulting in a depolarization by allowing positive charges to flow from the extracellular space across the neuronal membrane; this is followed by the delayed opening of potassium channels (hence their name "delayed rectifiers") that elicits the exit of positive charges and contributes to the repolarization of the membrane. If the temporal overlap between the entry of sodium and the outflow of potassium is high, the system will be relatively inefficient, because a larger number of positive charges—that is, sodium ions—will have to enter the cell in order to achieve a given depolarization. This is because, in terms of positive charges, the exit of potassium will partially counterbalance the entry of sodium. On the contrary, a low temporal overlap will optimize conditions to achieve depolarization. Calculations based on the biophysical properties of neuronal membranes predict a minimum positive charge transfer (carried by sodium) per propagating action potential of 121 nanoCoulombs/cm^2.

In their classical modeling of the AP in the squid giant axon, Hodgkin and Huxley estimated that this minimum was to be 4 times higher in order to charge the membrane capacitance and elicit depolarization. In axons of the rat hippocampus, the sodium channel rapid activation and fast decay relative to the delayed onset of potassium currents result in optimal charge separation (i.e., minimal overlap of the inward and outward currents). Under these circumstances, charging the membrane capacitance requires 153 nanoCoulombs/cm^2, representing only 1.3 times the theoretical minimum. In previous energy budget estimates, a factor of 4 was used to approximate the sodium charge transfer per propagating AP (see Attwell & Iadecola, 2002), resulting in an overestimation of the actual number of picomoles of sodium ions flowing across the axonal membrane per surface unit per action potential. Since reestablishing intracellular sodium homeostasis is operated by the energy-consuming sodium-potassium ATPase, an overestimation of sodium transfer will result in an overestimation of the energy cost of APs by a factor of approximately 3 (4:1.3).

Comparison of the charge transfer due to excitatory PSPs mediated by the glutamate receptor activation indicates that the charge transfer involved in AP is about 6 times lower than that underlying PSPs—15 versus 96 picoCoulombs, respectively—with marginal contributions by backpropagating APs (3 pC) and depolarization of presynaptic terminals (1 pC). Translated into energy costs, these measurements indicate that PSPs dominate the energy requirements for neuronal signaling. Since glutamate synapses represent at least 80% of cortical synapses, these data support previous experimental results *in vivo* and *in vitro*, suggesting that glutamatergic neurotransmission is the major determinant of energy utilization by the brain (Hyder et al., 2006). As will be discussed in subsequent sections, coupling of glutamatergic transmission to local energy (glucose) delivery is mediated by glutamate signaling on astrocytes via glutamate transporters. This astrocyte-based coupling mechanism and the associated process of glutamate recycling represent only approximately 15% of the overall energy cost of glutamatergic neurotransmission (Hyder et al., 2006).

In addition to the maintenance of ionic gradients that are disrupted during activity, other energy-consuming processes exist in neurons. Thus, the permanent synthesis of molecules needed for communications, such as neurotransmitters, or for general cellular purposes consumes energy. Axonal transport of molecules synthesized in the nucleus to their final destination along the axon or at the axon terminal is yet another process fueled by cellular energy metabolism.

Summary

Exactly as in other tissues, the metabolism of glucose, the main energy substrate of the brain, produces two forms of energy: ATP and NADPH. Glycolysis and the TCA cycle produce ATP, whereas energy in the form of reducing equivalents stored in the NADPH molecule is produced predominantly through the pentose phosphate pathway. Reduced glutathione provides a major defense against oxidative stress. It is mainly synthetized in astrocytes, implying that neurons rely on astrocytes to provide them with GSH (through the shuttling of GSH) for protection against oxidative stress insults. Maintenance of the electrochemical gradients, particularly for Na^+ and K^+,

mainly dissipated by glutamate-mediated postsynaptic potentials is the principal energy-consuming process of neural cells.

GLUTAMATE METABOLISM: A COORDINATED SHUTTLE BETWEEN ASTROCYTES AND NEURONS

Astrocytes are known to play a major role in numerous functions of the central nervous system, including glutamate, ion and water homeostasis, defense against oxidative stress, energy storage in the form of glycogen, scar formation and tissue repair, modulation of synaptic activity via the release of gliotransmitters, and synapse formation and remodeling. Importantly, essential astrocytic functions involve strong metabolic cooperations with neighboring neurons. Such a metabolic interaction has been reviewed earlier in this chapter for GSH metabolism but is also operative for K^+ and pH buffering, ascorbic acid recycling, energy substrate delivery, and glutamate metabolism in the so-called glutamate-glutamine shuttle (Allaman et al., 2011; Bélanger & Magistretti, 2009) (Fig. 12. 8). The latter represents a key function of astrocytes, which play a critical role in the regulation of glutamate neurotransmission.

Indeed, synaptically released glutamate is removed rapidly from the extracellular space by a transporter-mediated reuptake system that is particularly efficient in astrocytes. This mechanism contributes in a crucial manner to the fidelity of glutamate-mediated neurotransmission. Indeed, glutamate levels in the extracellular space are low (<3 μM), allowing for optimal glutamate-mediated signaling after depolarization while preventing overactivation of glutamate receptors, which could eventually result in excitotoxic neuronal damage.

Five glutamate transporter subtypes have been cloned in various species, including humans. The excitatory amino acid transporter 1 (EAAT 1) and EAAT 2 subtypes (respectively GLAST and GLT-1 in rodents) are localized exclusively in astrocytes, whereas the EAAT 3 and 4 are predominantly localized in neurons, with a widespread distribution for EAAT 3 and a localized distribution to Purkinje neurons for EAAT 4. EAAT 5 is localized in rod photoreceptors and in bipolar cells of the retina. The density of EAAT 1 and 2 is particularly high on astrocytes that surround axon terminals and dendritic spines, consistent with the prominent role of these transporters in the reuptake of synaptically released glutamate. The driving force for glutamate uptake through the specific transporters is the transmembrane Na^+ gradient; indeed, glutamate is cotransported with Na^+ with a stoichiometry of one glutamate for three Na^+ ions. The selective loss of EAAT 2, the astrocyte-selective glutamate transporter, has been demonstrated in the motor cortex and spinal cord of patients who died of Amyotrophic Lateral Sclerosis, a neurodegenerative disease affecting motor neurons.

One may wonder how astrocytes dispose of the glutamate that they take up, because, unlike carbohydrates or lipids, amino acids cannot be stored. The predominant pathway in peripheral tissues for disposing of amino acids is the transfer of their α amino group to a corresponding α-keto acid; this reaction is catalyzed by aminotransferases (Fig. 12.9). In astrocytes, the α amino group of glutamate can be transferred to oxaloacetate to yield α-ketoglutarate (α-KG) and aspartate in a reaction catalyzed by aspartate amino transferase (AAT). The α-KG generated is an intermediate of the TCA cycle and is therefore oxidized further. Another transamination reaction catalyzed by alanine amino transferase (ALAT) transfers the α amino group of glutamate to pyruvate, resulting in the formation of alanine and α-KG.

Two other pathways exist in astrocytes to metabolize glutamate. First, glutamate can be converted directly into α-KG through an NAD-requiring oxidative deamination catalyzed by glutamate dehydrogenase (GDH) (Fig. 12.9). Glutamate, by entering the TCA cycle indirectly (through AAT or ALAT) or directly (through GDH), is an energy substrate for astrocytes. Second, the quantitatively predominant metabolic pathway of glutamate in astrocytes is its amination to glutamine, an ATP-requiring reaction in which an ammonium ion is fixed on glutamate (Figs. 12.9 and 12.10). This reaction is catalyzed by glutamine synthase (GS), an enzyme almost exclusively localized in astrocytes, and provides an efficient means of disposing not only of glutamate but also of ammonium (see Box 12.2). Glutamine is released by astrocytes and is taken up by neurons, where it is hydrolyzed back to glutamate by the phosphate-dependent

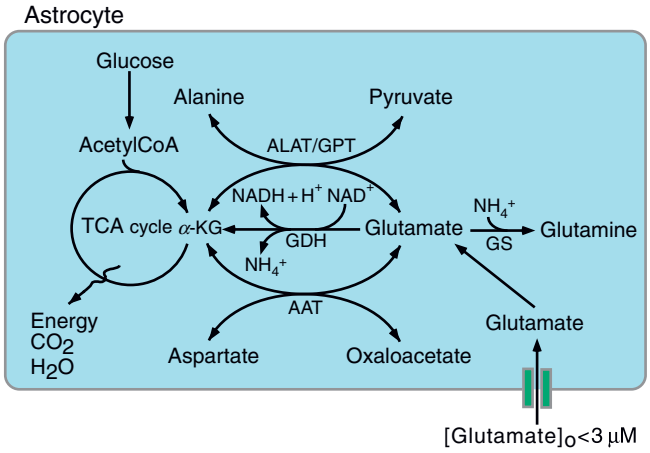


FIGURE 12.9 Metabolic fate of glutamate taken up by astrocytes. ALAT, alanine aminotransferase; GDH, glutamate dehydrogenase; GS, glutamine synthase; AAT, aspartate aminotransferase; GPT, glutamate dehydrogenase; α-KG, α-ketoglutarate.

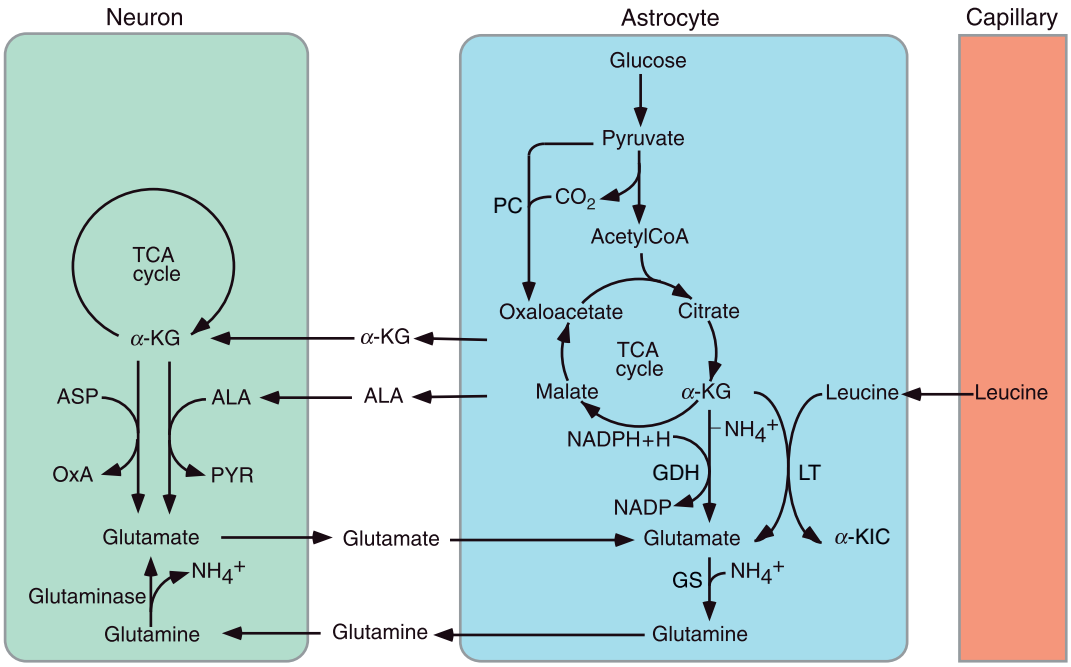


FIGURE 12.10 Metabolic intermediates are released by astrocytes to regenerate the glutamate neurotransmitter pool in neurons. Glutamine, formed from glutamate in a reaction catalyzed by glutamine synthase (GS), is released by astrocytes and taken up by neurons, which convert it into glutamate under the action of glutaminase. GS is an enzyme selectively localized in astrocytes. This metabolic cycle is referred to as the glutamate-glutamine shuttle. Other quantitatively less important sources of neuronal glutamate are alanine (ALA) and α-ketoglutarate (α-KG). In astrocytes, glutamate is synthesized *de novo* from α-KG through two different reactions catalyzed i) by glutamate dehydrogenase (GDH) and ii) by leucine transaminase (LT) producing glutamate and α-ketoisocaproate (α-KIC), in a transamination reaction in which leucine provides the amino group. Carbons "lost" from the TCA cycle as α-KG is converted to glutamate are replenished by oxaloacetate (OxA) formed from pyruvate (PYR) in a reaction catalyzed by pyruvate carboxylase (PC), another astrocyte-specific enzyme (anaplerotic cycle). ASP, aspartate.

BOX 12.2

HEPATIC ENCEPHALOPATHY

Hepatic encephalopathy is observed in patients with severe liver failure. The disease can be in one of two forms: an acute form, called fulminant hepatic failure, and a chronic form, called portosystemic encephalopathy. The neuropsychiatric symptoms of fulminant hepatic failure are delirium, coma, and seizures associated with acute toxic or viral hepatic failure. Patients having portosystemic encephalopathy may present personality changes, episodic confusion, or stupor and, in the most severe cases, coma. The current view on the pathophysiology of hepatic encephalopathy is that, due to liver failure, "toxic" substances that affect brain function accumulate in the circulation.

One of the substances thought to be responsible for neuropsychiatric "toxicity" is ammonia. The neuropathological findings are rather striking: astrocytes are the brain cells that appear to be principally affected. In the acute form, astrocyte swelling is prominent and likely to be the cause of the observed acute brain edema. In portosystemic encephalopathy, astrocytes adopt characteristic morphological features: in these cells, the nucleus is pale and enlarged, chromatin is marginated, and a prominent nucleolus is often observed. Lipofuscin deposits may be present, and the amount of the astrocyte-specific protein glial fibrillary acidic protein is decreased. Neurons appear structurally normal. All the foregoing histopathological changes have been reproduced *in vitro* by acutely or chronically applying ammonium chloride to primary astrocyte cultures. As mentioned earlier, detoxification of ammonium is an ATP-requiring, astrocyte-specific reaction catalyzed by glutamine synthase (see Figs. 12.9 and 12.10). It is therefore not surprising that excess ammonia perturbs energy metabolism; indeed, ammonia stimulates glycolysis, whereas it inhibits TCA cycle activity and decreases the glycogen content of astrocytes markedly. In addition, astrocytic accumulation of the osmotically active

BOX 12.2 *(cont'd)*

glutamine as a result of ammonia detoxification is thought to contribute at least partially to the swelling of astrocytes in hyperamonemic conditions.

In summary, while the precise pathophysiological mechanisms of the neuropsychiatric syndrome in hepatic encephalopathy are still unknown, this clinical condition provides a striking illustration of the fundamental importance of neuron–astrocyte metabolic interactions because structural and functional alterations apparently restricted to astrocytes result in severe behavioral perturbations.

enzyme glutaminase. This metabolic pathway is often referred to as the glutamate-glutamine shuttle (Figs. 12.8 and 12.10). It allows the removal of potentially toxic excess glutamate from the extracellular space, while returning to the neuron a synaptically inert (glutamine does not affect neurotransmission) precursor with which to regenerate the neuronal pool of glutamate.

However, not all glutamate is regenerated through the glutamate-glutamine shuttle because some of the glutamate released by neurons enters at the α-KG level, the TCA cycle in astrocytes; therefore, *de novo* synthesis is required to maintain the neuronal glutamate pool. Different pathways exist to synthesize glutamate both in astrocytes and neurons. In astrocytes, glutamate can be synthesized through amination of α-KG by two different pathways: GDH and leucine transaminase (LT) using leucine as a nitrogen donor (Fig. 12.10). Since these reactions take place in astrocytes, to replenish the neuronal pool, the astrocytes export glutamate as glutamine. Alternatively, neurons can take up α-KG and alanine from the extracellular space and convert them to glutamate and pyruvate in a transamination reaction catalyzed by ALAT (Fig. 12.10). In this case, as for the glutamate-glutamine shuttle, astrocytes provide the substrate(s) necessary for glutamate synthesis in neurons, since they are known to release significant amount of alanine an α-KG.

Note that because α-KG is used for glutamate synthesis, metabolic intermediates downstream of α-KG must be available to maintain a sustained flux through the TCA cycle in astrocytes (Fig. 12.10). This need is met by the activity of the enzyme pyruvate carboxylase (PC), which fixes CO_2 on pyruvate to generate oxaloacetate, which, by condensing with acetyl-CoA, maintains the flux through the TCA cycle. The carboxylation of pyruvate to oxaloacetate is referred to as an *anaplerotic* (Greek for "fill up") reaction. Interestingly, like glutamine synthase, PC is selectively localized in astrocytes. The fact that these two enzymes are localized in astrocytes in conjunction with the existence of a glutamate-glutamine shuttle stresses the essential role of astrocytes for maintaining the neuronal glutamate pool used for neurotransmission (Fig. 12.10).

Summary

A key function of astrocytes is to remove synaptically released glutamate. A large proportion of glutamate is transformed to glutamine through an energy-requiring process that also allows for the detoxification of ammonium. Glutamine released by astrocytes regenerates the neuronal glutamate pool. Some of the glutamate is also regenerated from astrocytic-derived metabolites either in astrocytes or neurons providing another indication of the crucial function that astrocytes play in maintaining the neuronal glutamate pool at levels that ensure the maintenance of synaptic transmission.

BRAIN ENERGY METABOLISM AT THE CELLULAR LEVEL

Glia and Vascular Endothelial Cells, in Addition to Neurons, Contribute to Brain Energy Metabolism

Neurons exist in a variety of sizes and shapes and express a large spectrum of firing properties (see Chapter 3). These differences are likely to imply specific energy demands; for example, large pyramidal cells in the primary motor cortex, which must maintain energy-consuming processes such as ion pumping over a large membrane surface or axonal transport along several centimeters, are likely to have considerably larger energy requirements than local interneurons. However, it is now clear that other cell types of the nervous system—glia and vascular endothelial cells—not only consume energy but also play a crucial role in the flux of energy substrates to neurons. Arguments for such an active role for nonneuronal cells—in particular, glia—are both quantitative and qualitative. Glial cells make up approximately half of the brain volume. A conservative figure is a 1:1 ratio between the number of astrocytes, one of the predominant glial cell types (see Chapters 1 and 3), and neurons. Higher ratios have been described, depending on the regions, developmental ages, or species. Indeed, the astrocyte-to-neuron ratio increases with the size of the brain and is thus high in humans. It is therefore

clear that glucose reaching the brain parenchyma provides energy substrates to a variety of cell types, only some of which are neurons.

Even more compelling for the realization of the key role that astrocytes play in providing energy substrates to active neurons are the cytological relations that exist among brain capillaries, astrocytes, and neurons. These relations, which are illustrated in Figure 12.11, are as follows. First, through specialized processes, called end feet, astrocytes surround brain capillaries. This implies that astrocytes contribute to the first cellular barrier that glucose entering the brain parenchyma encounters and make them a likely site of prevalent glucose uptake and energy substrate distribution. More than a century ago, the Italian histologist Camillo Golgi and his pupil Luigi Sala sketched such a principle. In addition to perivascular end feet, astrocytes bear processes that ensheathe synaptic contacts and express receptors and uptake sites with which neurotransmitters released during synaptic activity can interact (in particular glutamate, see previous sections) (Chapters 6 and 7). These features endow astrocytes with an exquisite sensitivity to detect increases in synaptic activity. In addition to this, and as previously described, astrocytes establish with neurons numerous metabolic cooperations that play a central role in brain homeostasis (Fig. 12.8). In summary, because of the foregoing structural and functional characteristics, astrocytes are ideally suited to couple local changes in neuronal activity with coordinated adaptations in energy metabolism (Fig. 12.11).

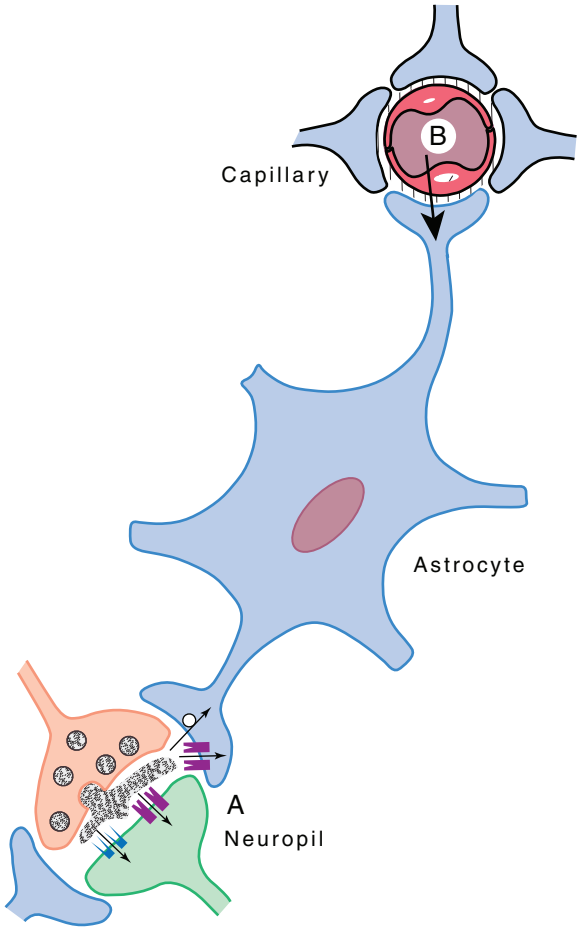


FIGURE 12.11 Schematic representation of cytological relations existing among intraparenchymal capillaries, astrocytes, and the neuropil. Astrocyte processes surround capillaries (end feet) and ensheathe synapses. In addition, receptors and uptake sites for neurotransmitters are present on astrocytes. These features make astrocytes ideally suited to sense synaptic activity (A) and to couple it with uptake and metabolism of energy substrates originating from the circulation (B).

A Tightly Regulated Glucose Metabolism Occurs in All Cell Types of the Brain, Neuronal and Nonneuronal

Given the high degree of cellular heterogeneity of the brain, understanding the relative role played by each cell type in the flux of energy substrates has largely depended on the availability of purified preparations, such as primary cultures enriched in neurons, astrocytes, or vascular endothelial cells. Such preparations have some drawbacks because they may not necessarily express all the properties of the cells *in situ.* In addition, one of the parameters of energy metabolism *in vivo*—namely, blood flow—cannot be examined in cultures. Despite these limitations, *in vitro* studies in primary cultures have proved very useful in identifying the cellular sites of glucose uptake and its subsequent metabolic fate—particularly, glycolysis and oxidative phosphorylation—thus providing illuminating correlations of two parameters of brain energy metabolism that are monitored *in vivo*: glucose utilization and oxygen consumption.

Glucose Transporters in the Brain

Glucose is a highly hydrophilic molecule that enters cells through a facilitated transport mediated by specific transporters. Twelve genes, encoding glucose transporter proteins, have been identified and cloned so far; these are designated GLUT1 to GLUT12. Glucose transporters belong to a family of rather homologous glycosylated membrane proteins with 12 transmembrane-spanning domains, and both amino and carboxyl terminals are exposed to the cytoplasmic surface of the membrane. In the brain, seven transporters are expressed predominantly in a cell-specific manner: GLUT1 (two isoforms), GLUT2-5, and GLUT8 (Fig. 12.12).

GLUT 1 and 3 represent the two major cerebral GLUTs. Two isoforms of GLUT1 with molecular masses of 55 and 45 kDa, respectively, are detected in the brain, depending on their degree of glycosylation. The 55-kDa form of

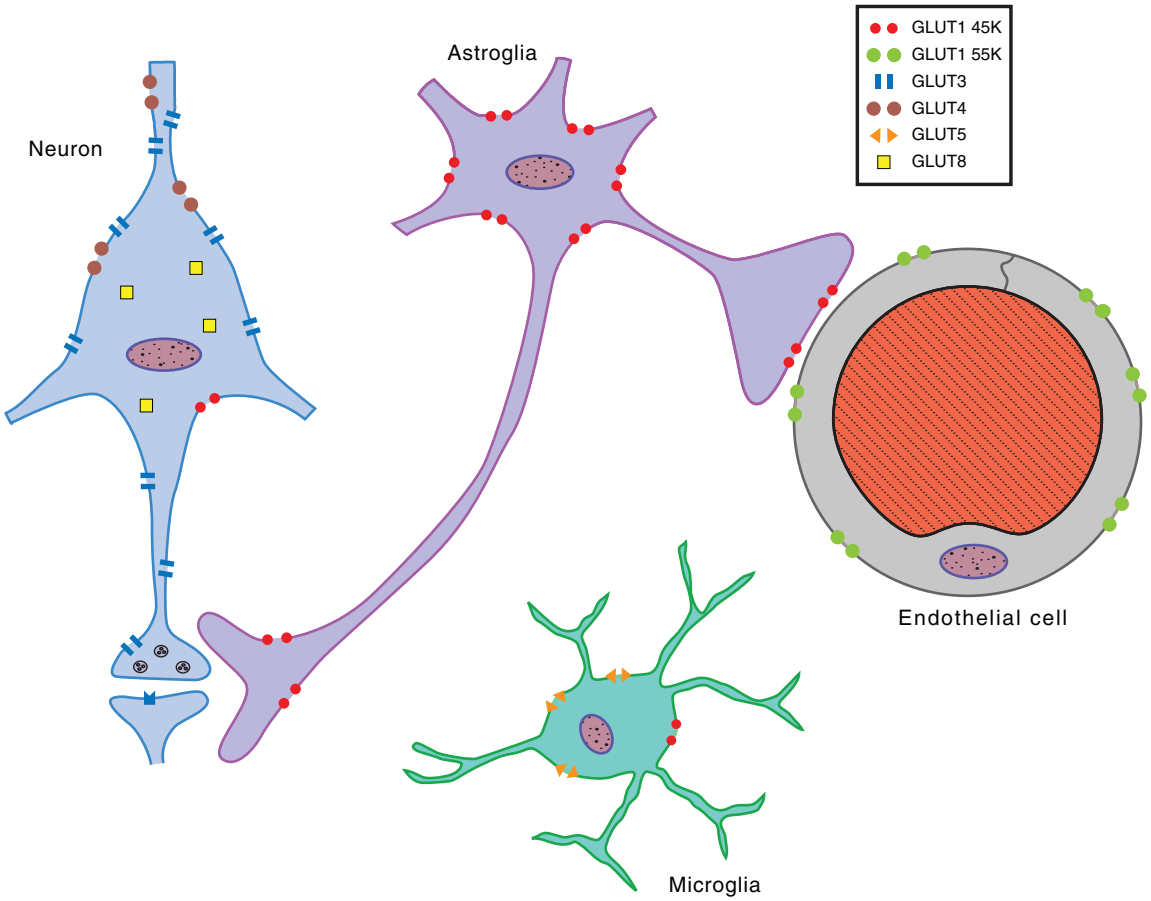


FIGURE 12.12 Cellular distribution of the principal glucose transporters in the nervous system. The 55-kDa form of GLUT1 is localized in brain microvessels, choroid plexus, and ependymal cells, whereas the 45-KDa form is localized predominantly in astrocytes. GLUT3 is specifically expressed in neurons. GLUT 4 and 8 are expressed in a subset of neurons, where their expression appears to predominate on the cell body and proximal dendrites. GLUT5 (which presents a higher affinity for fructose than glucose) is localized selectively to microglial cells. Another glucose transporter, GLUT2, has been localized selectively in astrocytes of discrete brain areas, such as certain hypothalamic and brainstem nuclei, which participate in the regulation of feeding behavior and in the central control of insulin release (not shown).

GLUT1 is essentially localized in brain microvessels, choroid plexus, and ependymal cells. In the brain *in situ*, the 45-kDa form of GLUT1 is localized predominantly in astrocytes. Under culture conditions, all neural cells, including neurons and other glial cells, can express GLUT1; however, this phenomenon appears to be due to the capacity of GLUT1 to be induced by cellular stress. As opposed to GLUT1, the glucose transporter specific for neurons is GLUT3; it is widely expressed across the different neuronal types. Its cellular distribution appears to predominate in the neuropil.

It is clear that glucose uptake into the brain parenchyma is a highly specified process regulated in a cell-specific manner by glucose transporter subtypes. Figure 12.12 summarizes this process: Glucose enters the brain through 55-kDa GLUT1 transporters localized on endothelial cells of the blood-brain barrier, thus representing the first barrier for the transport of circulating glucose into the brain parenchyma. Uptake into astrocytes is mediated by 45-kDa GLUT1 transporters, whereas GLUT3 transporters mediate this process in neurons.

Cell-Specific Glucose Uptake and Metabolism

With respect to their different physiological roles, astrocytes and neurons express different metabolic phenotypes (Fig. 12.8). As previously reviewed, specific functions of astrocytes, in particular with regard to astrocyte-neuron metabolic interactions, rely on the selective expression of key proteins (e.g., PC, EAAT1 and 2 or GS). Such differences in metabolic profiles also operate for glucose metabolism. Indeed, cell type–specific expression patterns of key genes regulating energy metabolism have been shown (see, for example, Bélanger et al., 2011, and references therein) for glucose (see above) and MCTs, for genes involved in the regulation of the transfer of NADH into the mitochondria (i.e., malate-aspartate shuttle and glycerol phosphate shuttle), and for metabolic fluxes of

glucose, including mitochondrial activity, of glycogen synthesis or glycolysis. For example, glucose storage in the form of glycogen is exclusively found in astrocytes (see below), in part because the glycogen synthesis machinery is kept inactive in neurons through proteasomal-dependent mechanisms (see Bélanger et al., 2011). Adaptation of energy metabolism to increased energy demand also presents differences between the two cell types. For instance, in a series of elegant studies Moncada, Bolanos, and their collaborators observed that, in contrast to astrocytes, neurons are unable to prevent nitrosative stress damages through upregulation of glycolysis due to the virtual absence of a key regulator of glycolysis in this cell type. Further stressing the specificity in their metabolic profile, manipulations that drive neurons to increase glucose utilization or to synthetize glycogen affect their survival (see Bélanger et al., 2011). In summary, neurons and astrocytes preferentially use different (but complementary) metabolic pathways with regard to glucose metabolism under physiological conditions.

As we have seen, glucose utilization can be assessed with radioactively labeled 2-DG. To determine the cellular site of basal and activity-related glucose utilization, this technique has been applied to homogeneous cultures of astrocytes or neurons. For quantitative purposes and to allow comparisons with *in vivo* studies, these *in vitro* experiments, in which radioactive 2-DG is used as a tracer, must be conducted in a medium containing a concentration of glucose near that measured *in vivo* in the extracellular space of the brain (0.5 to 2 mM). The basal rate of glucose utilization is higher in astrocytes than in neurons, with values of about 20 and 6 nmol per milligram of protein per minute, respectively (Magistretti & Pellerin, 1999). These values are of the same order as those determined *in vivo* for cortical gray matter (10–20 nmol mg^{-1} min^{-1}) with the 2-DG autoradiographic technique. In view of this difference and of the quantitative preponderance of astrocytes compared with neurons in the gray matter, these data reveal a significant contribution by astrocytes to basal glucose utilization as determined by 2-DG autoradiography or PET *in vivo.* High-resolution microautoradiographic imaging has indicated an approximately even distribution of 2-DG in neurons and astrocytes.

The contribution of astrocytes to glucose utilization during activation appears to be even more striking. *In vitro,* activation can be mimicked by exposure of the cells to glutamate, the principal excitatory neurotransmitter (Chapter 6), because, during activation of a given cortical area, the concentration of glutamate in the extracellular space increases considerably due to its release from the axon terminals of activated pathways. L-glutamate stimulates 2-DG uptake and phosphorylation by astrocytes in a concentration-dependent manner, with an EC_{50} of 60 to 80 μM (Pellerin & Magistretti, 1994). Unlike other actions of glutamate, stimulation of glucose utilization in astrocytes is mediated not by specific glutamate receptors, but by glutamate transporters as well. As previously described, glutamate uptake into astrocytes is mediated by cell specific transporters using transmembrane Na^+ gradient as a driving force (Fig. 12.8).

Glutamate-Stimulated Uptake of Glucose by Astrocytes Is a Source of Insight into the Cellular Bases of ^{18}F–2–DG PET *in Vivo*

The glutamate-stimulated uptake of glucose by astrocytes is a source of insight into the cellular bases of the activation-induced local increase in glucose utilization visualized with ^{18}F-2-DG PET *in vivo.* As we have seen, focal physiological activation of specific brain areas is accompanied by increases in glucose utilization; because glutamate is released from excitatory synapses when neuronal pathways subserving specific modalities are activated, the stimulation by glutamate of glucose utilization in astrocytes provides a direct mechanism for coupling neuronal activity to glucose utilization in the brain (Fig. 12.13). The intracellular molecular mechanism of this coupling requires Na^+, K^+-ATPase because ouabain completely inhibits the glutamate-evoked 2-DG uptake by astrocytes (Pellerin & Magistretti, 1994). The astrocytic Na^+, K^+-ATPase responds predominantly to increases in intracellular Na^+ (Na^+_i), for which it shows a K_m of about 10 mM. In astrocytes, the Na^+_i concentration ranges between 10 and 20 mM, so Na^+, K^+-ATPase is set to be activated readily when Na^+_i rises concomitantly with glutamate uptake (Magistretti & Chatton, 2005) (Fig. 12.14). These observations indicate that a major determinant of glucose utilization is the activity of Na^+, K^+-ATPase. In this context, one should note that, *in vivo*, the main mechanism that accounts for activation-induced 2-DG uptake is the activity of Na^+, K^+-ATPase. It is important here to briefly consider the relative participation of the neuronal and astrocytic Na^+, K^+-ATPases in glucose utilization. When glutamate is released from depolarized neuronal terminals, it is taken up predominantly into astrocytes. The stoichiometry of glutamate reuptake being one molecule of glutamate cotransported with three Na^+ ions, the increase in intracellular astrocytic Na^+ concentration associated with glutamate reuptake massively activates the pump. Thus, although the tonic activity of the Na^+, K^+-ATPase is needed to maintain the transmembrane neuronal and glial ionic gradients and accounts for basal glucose utilization, on a short-term temporal scale (from milliseconds to seconds), when glutamate is released from depolarized axon terminals of modality-specific afferents, the astrocytic Na^+, K^+-ATPase is briskly activated, due to the massive increase (by at least 10 mM) in intracellular Na^+ associated with glutamate reuptake, providing the signal for the activation-dependent glucose utilization.

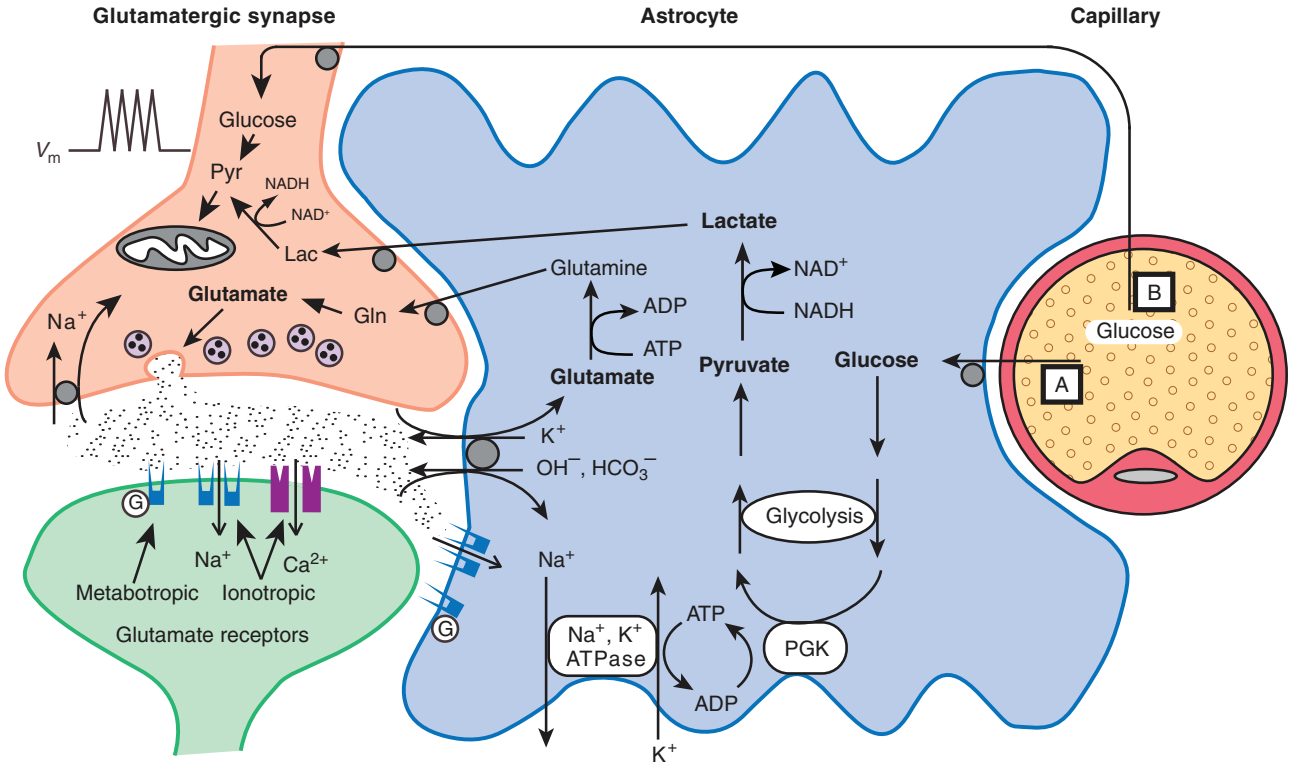


FIGURE 12.13 Schematic representation of the mechanism for glutamate-induced glycolysis in astrocytes during physiological activation. At glutamatergic synapses, presynaptically released glutamate depolarizes postsynaptic neurons by acting at specific receptor subtypes. The action of glutamate is terminated by an efficient glutamate uptake system located primarily in astrocytes. In fact, glutamate is taken up predominantly into astrocytes. Glutamate is cotransported with Na^+, resulting in an increase in the intra-astrocytic concentration of Na^+, leading to an activation of the astrocytic Na^+, K^+-ATPase. Activation of Na^+,K^+-ATPase stimulates glycolysis (i.e., glucose utilization and lactate production). The stoichiometry of this process is such that for one glutamate molecule taken up with three Na^+ ions, one glucose molecule enters astrocytes, two ATP molecules are produced through glycolysis, and two lactate molecules are released. Within the astrocyte, one ATP fuels one "turn of the pump," while the other provides the energy needed to convert glutamate to glutamine by glutamine synthase (see Fig. 12.10). Once released by astrocytes, lactate can be taken up by neurons and serve as an energy substrate (For graphic clarity, only lactate uptake into presynaptic terminals is indicated. However, this process could also take place at the postsynaptic neuron). This model, which summarizes *in vitro* experimental evidence indicating glutamate-induced glycolysis, is taken to show cellular and molecular events occurring during activation of a given cortical area (arrow labeled A, activation). Direct glucose uptake into neurons under basal conditions is also shown (arrow labeled B, basal conditions). Pyr, pyruvate; Lac, lactate; Gln, glutamine; G, G protein; PGK, phosphoglycerate kinase. *Modified with permission from Pellerin and Magistretti (1994).*

How does activation of Na^+, K^+-ATPase cause increased glucose utilization? The mechanism was explained by pioneering studies on erythrocytes by Joseph Hoffmann and colleagues at Yale University, which have been confirmed in a number of other cell systems, including brain and vascular smooth muscle. The increase in pump activity consumes ATP (Fig. 12.14), which is a negative modulator of phosphofructokinase, the principal rate-limiting enzyme of glycolysis (Fig. 12.1). Thus, when ATP concentration is low, phosphofructokinase activity is stimulated, resulting in increased glucose utilization. The activity of hexokinase, the enzyme responsible for glucose and 2-DG phosphorylation (Fig. 12.5), is also increased under these conditions. This explains why the increase in glucose utilization, associated with the stimulation of Na^+,K^+-ATPase, can be monitored with 2-DG, which is not processed beyond the hexokinase step.

A compartmentalization of glucose uptake during activation has also been unequivocably found by Marco Tsacopoulos and colleagues in the honeybee drone retina (Tsacopoulos & Magistretti, 1996). In this highly organized, crystal-like nervous tissue preparation, photoreceptor cells form rosette-like structures that are surrounded by glial cells. In addition, mitochondria are exclusively present in the photoreceptor neurons. Light activation reveals an increase in radioactive 2-DG uptake in the glial cells surrounding the rosettes but not in the photoreceptor neurons. An increase in O_2 consumption is nevertheless measured in photoreceptor neurons. After activation of photoreceptors by light, glucose is probably taken up predominantly by glial cells, which then release a metabolic substrate to be oxidized by photoreceptor neurons.

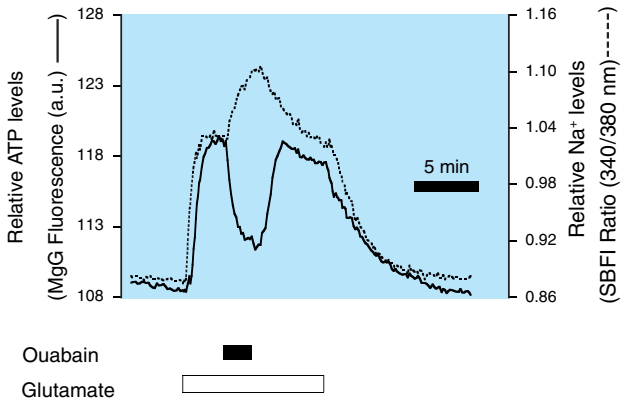


FIGURE 12.14 Temporal coincidence in the increase in sodium concentration and ATP consumption triggered by glutamate in astrocytes. These processes, which are dependent on the activity of the Na^+,K^+-ATPase as they are ouabain-sensitive, are the effectors of the glutamate-stimulated glycolysis in astrocytes (see also Fig. 12.13). *Modified with permission from Magistretti and Chatton (2005).*

In summary, as indicated in the operational model described in Figure 12.13, upon activation of a particular brain area, glutamate released from excitatory terminals is taken up by a Na^+-dependent transporter located on astrocytes. The ensuing local increase in intracellular Na^+ concentration activates Na^+,K^+-ATPase, which in turn stimulates glucose uptake by astrocytes. The key role of glial glutamate transporters in activity-dependent glucose uptake by the brain has been demonstrated *in vivo* (see Bélanger et al., 2011, and references therein). This model delineates a simple mechanism for coupling synaptic activity to glucose utilization; in addition, it is consistent with the notion that the signals detected during physiological activation in humans with ^{18}F-2-DG PET and autoradiography in laboratory animals may predominantly reflect uptake of the tracer into astrocytes. Recently, in the well-studied whisker-to-barrel somatosensory pathway, Buzsaki and colleagues have demonstrated a selective increase in glucose uptake into astrocytes during activation. The model originally proposed by Pellerin and Magistretti (1994) does not question the validity of the 2-DG-based techniques; rather, it provides a cellular and molecular basis for these functional brain-imaging techniques (Magistretti & Pellerin, 1999).

Lactate Released by Astrocytes Is a Metabolic Substrate for Neurons

The fact that the increase in glucose uptake during activation can be ascribed predominantly, if not exclusively, to astrocytes indicates that energy substrates must be released by astrocytes to meet the energy demands of neurons. As indicated earlier, lactate and pyruvate are adequate substrates for brain tissue *in vitro* (Schurr, 2006). In fact, synaptic activity can be maintained *in vitro* in cerebral cortical slices with only lactate or pyruvate as a substrate. Lactate is quantitatively the main metabolic intermediate released by cultured astrocytes at a rate of 15 to 30 nmol per milligram of protein per minute. For lactate (or pyruvate) to be a metabolic substrate for neurons, particularly during activation, two additional conditions must be fulfilled: (1) that indeed during activation lactate release by astrocytes increases, and (2) lactate uptake by neurons must be demonstrated. Both mechanisms have been indeed demonstrated. Mimicking activation *in vitro* by exposing cultured astrocytes to glutamate results in a marked release of lactate and, to a lesser degree, pyruvate (Pellerin & Magistretti, 1994). This glutamate-evoked lactate release shows the same pharmacology and time course as glutamate-evoked glucose utilization and indicates that glutamate stimulates the processing of glucose through glycolysis. *In vivo* ^{1}H MRI studies in humans that show a transient lactate peak in different brain areas during physiological stimulation (Figley & Stroman, 2011, and references therein) are consistent with the notion of activation-induced transient glycolysis. Finally, MCTs have been demonstrated on neurons and astrocytes in addition to capillaries.

Thus, a metabolic compartmentation whereby glucose taken up by astrocytes and metabolized glycolytically to lactate is then released in the extracellular space to be utilized by neurons is consistent with biochemical and electrophysiological observations (Bélanger et al., 2011; Hyder et al., 2006; Tsacopoulos & Magistretti, 1996). This array of *in vitro* and *in vivo* experimental evidence, known as the Astrocyte-Neuron Lactate Shuttle (ANLS), is summarized in the model of cell-specific metabolic regulation illustrated in Figure 12.13.

Studies of the well-compartmentalized honeybee drone retina and of isolated preparations of guinea pig retina containing photoreceptors attached to Mueller (glial) cells corroborate the existence of such metabolic fluxes between glia and neurons. In addition to the glial localization of glucose uptake during activation, glycolytic products have been shown to be released. In particular, during activation, glial cells in the honeybee drone retina release alanine produced from pyruvate by transamination; the released alanine is taken up by photoreceptor neurons and, after reconversion into pyruvate, can enter the TCA cycle to yield ATP through oxidative phosphorylation (Fig. 12.2). In the guinea pig retina, lactate, formed glycolytically from glucose, is released by Mueller cells to fuel photoreceptor neurons (Tsacopoulos & Magistretti, 1996).

In addition to plasma lactate, which is imported into the brain to a limited extent or during specific conditions (see above), lactate formed *within* the brain parenchyma (e.g., through glutamate-activated glycolysis in astrocytes) can thus fulfill the energetic needs of neurons. Lactate, after conversion into pyruvate by a reaction catalyzed by lactate dehydrogenase (LDH), can provide, on a molar

basis, 14 to 17 ATPs through oxidative phosphorylation. Conversion of lactate to pyruvate does not require ATP, and, in this regard, lactate is energetically more favorable than the first obligatory step of glycolysis in which glucose is phosphorylated to glucose 6-phosphate at the expense of one molecule of ATP (Fig. 12.1). In addition, lactate may contribute to the redox potential of neurons, since through its conversion to pyruvate, it generates NADH, hence providing energy in the form of reducing equivalents useful for a variety of metabolic reactions but also for ROS scavenging after its conversion to NADPH (Fig. 12.7). Another metabolic fate for lactate has been shown *in vitro* and *in vivo* by MRS. Thus, once converted to pyruvate in neurons, lactate may enzymatically yield glutamate and thus be a substrate for the replenishment of the neuronal pool of glutamate. Because this reaction is not associated with oxygen consumption, part of the uncoupling between glucose utilization and oxygen consumption described in certain paradigms of activation may be explained by the processing of glucose-derived lactate into the glutamate neuronal pool; another reason for such an uncoupling may be the use of glucose for non-energy-yielding processes such as its incorporation into certain macromolecules (glycoproteins and glycolipids).

Very recent evidence demonstrates that in addition to its role as an energy substrate, astrocytic-derived lactate also participates in the control of neuronal activity. For instance, effects of lactate on neuronal excitability have been observed for glucose-sensing and orexin neurons in the hypothalamus, as well as GABAergic neurons in the subfornical organ (see Bélanger et al., 2011, and references therein). Although the mechanism(s) of action of lactate in this context are not yet fully characterized, they appear to critically involve ATP-sensitive K^+ channels (K_{ATP}) (see Bélanger et al., 2011, and references therein). In this case, increased ATP concentration, produced by lactate oxidation, would trigger the closure of K_{ATP} channels leading to membrane depolarization and neuronal activation, providing a mechanism that links lactate release by astrocytes, lactate utilization by neurons, and modulation of neuronal excitability (see Chapter 5). These observations add another level of interaction between astrocytes and neurons with regards to energy metabolism.

Glycogen, the Storage Form of Glucose, Is Localized in Astrocytes

Glycogen is the single largest energy reserve of the brain; it is mainly localized in astrocytes, although ependymal and choroid plexus cells, as well as certain large neurons in the brainstem, contain glycogen. When compared to the contents in liver and muscle, the glycogen concentration in the brain is exceedingly small, about 100 and 10 times inferior, respectively.

Glycogen Metabolism Is Coupled to Neuronal Activity

Glycogen turnover in the brain is extremely rapid, and glycogen levels are finely coordinated with synaptic activity. For example, during general anesthesia, a condition in which synaptic activity is markedly attenuated, glycogen levels rise sharply. Interestingly, however, the glycogen content of cultures containing exclusively astrocytes is not increased by general anesthetics; this observation indicates that the *in vivo* action of general anesthetics on astrocyte glycogen is due to the inhibition of neuronal activity, stressing the existence of a tight coupling between synaptic activity and astrocyte glycogen. Accordingly, reactive astrocytes, which develop in areas where neuronal activity is decreased or absent as a consequence of injury, contain high amounts of glycogen.

Certain Neurotransmitters Regulate Glycogen Metabolism in Astrocytes

Glycogen levels in astrocytes are tightly regulated by various neurotransmitters. Several monoamine neurotransmitters—namely, noradrenaline, serotonin, and histamine—are glycogenolytic in the brain, in addition to certain peptides, such as VIP and pituitary adenylate cyclase activating peptide (PACAP), and adenosine and ATP. The effects of all these neurotransmitters are mediated by their cogent specific receptors coupled to second messenger pathways that are under the control of adenylate cyclase or phospholipase C. These observations show that neuronal signals (e.g., certain neurotransmitters) can exert receptor-mediated metabolic effects on astrocytes in a manner similar to peripheral hormones on their target cells. However, the action of this type of neurotransmitter is temporally specified and spatially restricted to activated areas. Indeed, brain glycogenolysis in laboratory animals has also been demonstrated *in vivo* after physiological activation of modality-specific pathways (see Bélanger et al., 2011, and references therein). These observations indicate that the physiological activation of specific neuronal circuits results in the mobilization of glial glycogen store.

The initial rate of glycogenolysis activated by VIP and noradrenaline measured in *in vitro* or *ex vivo* preparations is between 5 and 10 nmol per milligram of protein per minute, a value that is remarkably close to glucose utilization of the gray matter, as determined by the 2-DG auto-radiographic method. This correlation indicates that glycosyl units mobilized in response to glycogenolytic neurotransmitters can provide quantitatively adequate substrates for the energy demands of the brain parenchyma. While glycogen mobilization may also fulfill the astrocytes' own metabolic needs, glycogen breakdown typically results in lactate production and release in the

extracellular space. Consistent with its role as a neuronal energy substrate, glycogen-derived lactate was demonstrated to preserve axonal function in conditions of energy deprivation. It was also observed that astrocytic glycogen mobilization is required to sustain neuronal activity during intense stimulation in a mouse optic nerve preparation (see Bélanger et al., 2011, and references therein). Furthermore, recent data indicate a critical role of glycogen metabolism in higher brain functions. Thus, glycogen-derived lactate, and its transfer to neurons, is necessary for the establishment of long-term memory formation and for the maintenance of long-term potentiation (LTP) *in vivo* (see Bélanger et al., 2011).

Summary

Under basal conditions, glucose uptake and metabolism occur in every brain cell type. Glucose uptake is mediated by specific transporters that are distributed in a cell-specific manner. Astrocytes play a critical role in the utilization of glucose coupled to excitatory synaptic transmission. The molecular mechanisms of this coupling are stoichiometrically directed: for each synaptically released glutamate molecule taken up with three Na^+ ions by an astrocyte, one glucose molecule enters the same astrocyte, two ATP molecules are produced through glycolysis, and two lactate molecules are released and consumed by neurons to yield 14-17 ATPs through oxidative phosphorylation. Neuronal signals—for example, certain neurotransmitters—can exert receptor-mediated glycogenolysis in astrocytes in a manner similar to peripheral hormones on their target cells. However, this type of effect by neurotransmitters is temporally specified and spatially restricted within activated areas, possibly to provide additional energy substrates (i.e., lactate) in register with local increases in neuronal activity. This glycogen-derived lactate is critical for the expression of synaptic plasticity.

THE ASTROCYTE-NEURON METABOLIC UNIT

From a strictly energetic viewpoint, the brain can be seen as an almost exclusive glucose-processing machine producing H_2O and CO_2. However, the metabolism of glucose in the brain is specified temporally, spatially, and functionally. Thus, glucose metabolism increases with exquisite spatiotemporal precision in register with neuronal activity. The site of this increase is not the neuronal cell body; rather, it is the neuropil, where presynaptic terminals, postsynaptic elements, and astrocytes ensheathing synaptic contacts are localized. This cytological relation between astrocytes and neurons is also manifested by a functional metabolic partnership that operates as one of the major elements at the basis of neurovascular and neurometabolic coupling (Fig. 12.15). In response to a neuronal signal (glutamate), glucose is taken up by perivascular astrocytes (neurometabolic coupling) through a

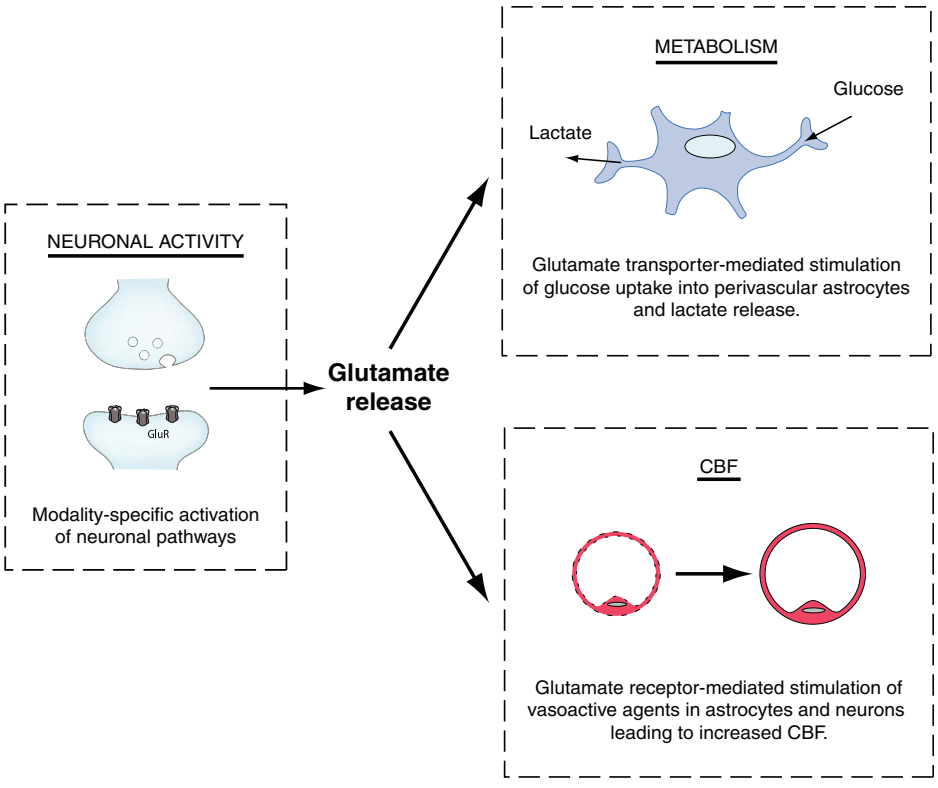

METABOLISM
Glucose
Lactate
Glutamate transporter-mediated stimulation of glucose uptake into perivascular astrocytes and lactate release.
NEURONAL ACTIVITY
GluR
Modality-specific activation of neuronal pathways
Glutamate release
CBF
Glutamate receptor-mediated stimulation of vasoactive agents in astrocytes and neurons leading to increased CBF.

FIGURE 12.15 Coupling "in-parallel" of neuronal activity with CBF and energy metabolism (i.e., glucose use). See text for discussion. GluR: Glutamate receptor. *Modified from Magistretti and Pellerin (1999).*

glutamate transporter-mediated mechanism and released as a glucose-derived metabolic substrate for neurons (lactate), and vasoactive agents are released by astrocytes and neurons to increase CBF (neurovascular coupling) involving receptor-mediated mechanisms (Fig. 12.15).

Glucose also provides the carbon backbone for regeneration of the neuronal pool of glutamate. This process results from a close astrocyte-neuron cooperation. Indeed, the selective localization of PC in astrocytes, indicating the need to replenish the TCA cycle with carbon backbones, strongly suggests that glucose-derived metabolic intermediates are used for glutamate (and other amino acid) synthesis. The newly synthesized glutamate is not provided as such by astrocytes to neurons; rather, it is converted to glutamine by GS, another enzyme localized selectively in astrocytes. Glutamate, taken up by astrocytes during synaptic activity, undergoes the same metabolic process, also being released as glutamine (the glutamate-glutamine shuttle).

Summary

In conclusion, the axon terminal of glutamatergic neurons, which are the main communication lines in the nervous system, and the astrocytic processes that surround them should be viewed as a metabolic unit in which the neuron furnishes the activation signal (glutamate) to the astrocyte and the astrocyte provides not only the precursors needed to maintain the neurotransmitter pool (i.e., mainly glutamine), but also the energy substrate (lactate) (Fig. 12.16). The efficacy of the predominant excitatory synapse in the brain, the glutamatergic synapse, cannot be maintained without a close astrocyte-neuron interaction. Furthermore, evidence is accumulating that indicates that dysfunctions of this neuron-astrocyte metabolic cooperation are likely to be involved in pathological conditions, in particular neurodegeneration (Allaman et al., 2011; Bélanger & Magistretti, 2009).

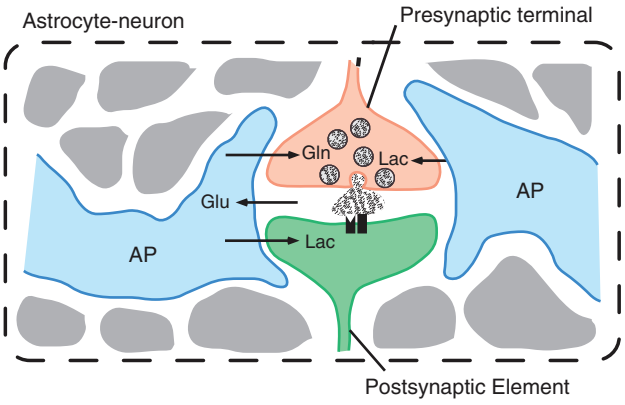


FIGURE 12.16 The astrocyte-neuron metabolic unit. Glutamatergic terminals and the astrocytic processes that surround them can be viewed as a highly specialized metabolic unit in which the activation signal (glutamate, Glu) is provided by the neuron to the astrocyte, whereas the astrocyte provides the precursors needed to maintain the neurotransmitter pool (i.e., mainly glutamine, Gln), as well as the energy substrate (lactate, Lac). AP, astrocyte process.

References

Allaman, I., Bélanger, M., & Magistretti, P. J. (2011). Astrocyte-neuron metabolic relationships: For better and for worse. *Trends in Neuroscience, 34,* 76–87.

Attwell, D., Buchan, A. M., Charpak, S., Lauritzen, M., MacVicar, B. A., & Newman, E. A. (2010). Glial and neuronal control of brain blood flow. *Nature, 468,* 232–243.

Attwell, D., & Iadecola, C. (2002). The neural basis of functional brain imaging signals. *Trends in Neuroscience, 25,* 621–625.

Bélanger, M., Allaman, I., & Magistretti, P. J. (2011). Brain energy metabolism: Focus on astrocyte-neuron metabolic cooperation. *Cell Metabolism, 14,* 724–738.

Bélanger, M., & Magistretti, P. J. (2009). The role of astroglia in neuroprotection. *Dialogues in Clinical Neuroscience, 11,* 281–295.

Cauli, B., Tong, X. K., Rancillac, A., Serluca, N., Lambolez, B., Rossier, J., & Hamel, E. (2004). Cortical GABA interneurons in neurovascular coupling: Relays for subcortical vasoactive pathways. *Journal of Neuroscience, 24,* 8940–8949.

Figley, C. R., & Stroman, P. W. (2011). The role(s) of astrocytes and astrocyte activity in neurometabolism, neurovascular coupling, and the production of functional neuroimaging signals. *European Journal Neuroscience, 33,* 577–588.

Frackowiak, R. S. J., Magistretti, P. J., Shulman, R. G., & Adams, M. (2001). *Neuroenergetics: Relevance for functional brain imaging*. Strasbourg: HFSP.

Hyder, F., Patel, A. B., Gjedde, A., Rothman, D. L., Behar, K. L., & Shulman, R. G. (2006). Neuronal-glial glucose oxidation and glutamatergic-GABAergic function. *Journal of Cerebral Blood Flow & Metabolism, 26,* 865–877.

Magistretti, P. J., & Chatton, J. Y. (2005). Relationship between L-glutamate-regulated intracellular Na(+) dynamics and ATP hydrolysis in astrocytes. *Journal of Neural Transmission, 112,* 77–85.

Magistretti, P. J., & Pellerin, L. (1999). Cellular mechanisms of brain energy metabolism and their relevance to functional brain imaging. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 354,* 1155–1163.

Pellerin, L., & Magistretti, P. J. (1994). Glutamate uptake into astrocytes stimulates aerobic glycolysis: A mechanism coupling neuronal activity to glucose utilization. *Proceedings of the National Academy of Sciences of the United States of America, 91,* 10625–10629.

Petzold, G. C., & Murthy, V. N. (2011). Role of astrocytes in neurovascular coupling. *Neuron, 71,* 782–797.

Raichle, M. E., & Mintun, M. A. (2006). Brain work and brain imaging. *Annual Review of Neuroscience, 29,* 449–476.

Schurr, A. (2006). Lactate: The ultimate cerebral oxidative energy substrate? *Journal of Cerebral Blood Flow and Metabolism, 26,* 142–152.

Tsacopoulos, M., & Magistretti, P. J. (1996). Metabolic coupling between glia and neurons. *Journal of Neuroscience, 16,* 877–885.

Igor Allaman and Pierre J. Magistretti

SECTION III

NERVOUS SYSTEM DEVELOPMENT

This page intentionally left blank

CHAPTER

13

Neural Induction and Pattern Formation

This chapter describes some of the key events that take place in the early stages of vertebrate nervous system development, during which the neural tube, placodes, and neural crest are formed, setting up the *structural ground plan* on which the functioning nervous system is built. The first part of the chapter describes how these embryonic structures are specified by inductive tissue interactions and how the processes of morphogenesis shape the early nervous system. The second part describes how the nervous system is patterned along its different axes, such that neural precursor cells acquire the ability to give rise to hundreds of different types of neuron, each forming in its appropriate position in the adult nervous system. These processes begin when neural tissue first appears and progress to finer and finer levels of cellular organization.

NEURAL INDUCTION

Embryonic Origins of the Nervous System

The progenitor cells that form the entire vertebrate nervous system can be traced back in development to an epithelial cell layer, the *ectoderm*, that covers the outside of the embryo during *gastrulation* (Fig. 13.1). Ectodermal cells give rise to different derivatives depending on position along the dorsoventral (DV) axis of the embryo. The dorsal-most ectoderm thickens to form the *neural plate*, a structure shaped like a tennis racquet with the head lying anteriorly. During a complex morphogenetic process called *neurulation*, the flat neural plate rolls up into a tube that sinks into the interior of the embryo and becomes overlain by epidermal ectoderm. This *neural tube* is the anlage of the central nervous system (CNS). As the neural plate folds and closes, *neural crest* cells detach from its lateral margins and migrate away, later condensing to form the major part of the peripheral nervous system (PNS). The ectodermal cells lying more ventral around the edge of the cranial neural plate constitute a domain where various sensory structures such as the ear, nose, and cranial sensory ganglia will arise from isolated patches of thickened ectoderm called *placodes*. Finally, ectodermal cells on the ventral side of the embryo give rise to the outer layer of the skin, the epidermis. The first step in forming the nervous system, therefore, is the establishment of these different regions of ectoderm, a process involving inductive interactions that were discovered early in the last century. Only quite recently, however, has the mechanism of *neural induction* been revealed at the molecular level.

Neural Induction and the Spemann Organizer

In the 1920s, Hilde Mangold, a PhD student in Hans Spemann's laboratory at the University of Freiburg, performed transplantation experiments on newt embryos and frequently obtained conjoined twins after grafting a small piece of the dorsal blastopore lip (DBL, where mesodermal cells start to involute during gastrulation) from a donor embryo to the opposite side of a host (Spemann & Mangold, 1924). Unlike others who had performed similar experiments before, Spemann and Mangold used differently pigmented newt species as donors and hosts, allowing them to determine the fate of the grafted cells. They noticed that the grafted DBL made only a small contribution to the secondary embryonic axis, most of which was derived from the host embryo, and concluded that DBL cells must have induced surrounding tissue in a spatially organized fashion to form an accurately patterned embryo (Fig. 13.2). Based on this inducing activity, Spemann called the DBL the *organizer*. For this groundbreaking discovery he was awarded with the 1935 Nobel Prize in Medicine (Mangold had died earlier in a domestic accident). Since

© 2013 Elsevier Inc. All rights reserved.

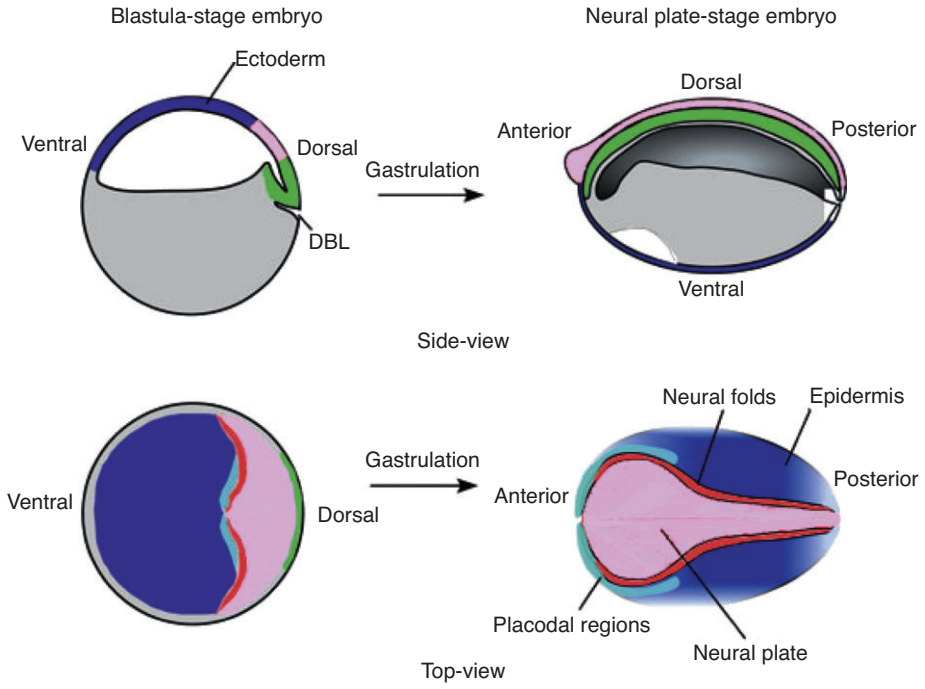


FIGURE 13.1 Ectoderm is subdivided into different fates during gastrulation. Shown is a cross-sectional side view, and top view of an amphibian embryo before (blastula-stage) and after gastrulation (neural plate-stage). During gastrulation, the gray and green (dorsal mesoderm) regions of the blastula involute, and the ectoderm (blue/red) at the top spreads over the outside of the embryo, and extends along the anterior-posterior axis. As this process occurs, signals emanating from the dorsal and ventral sides of the embryo specify different fates, so that dorsal ectoderm (red) becomes neural tissue while ventral ectoderm (blue) becomes epidermis.

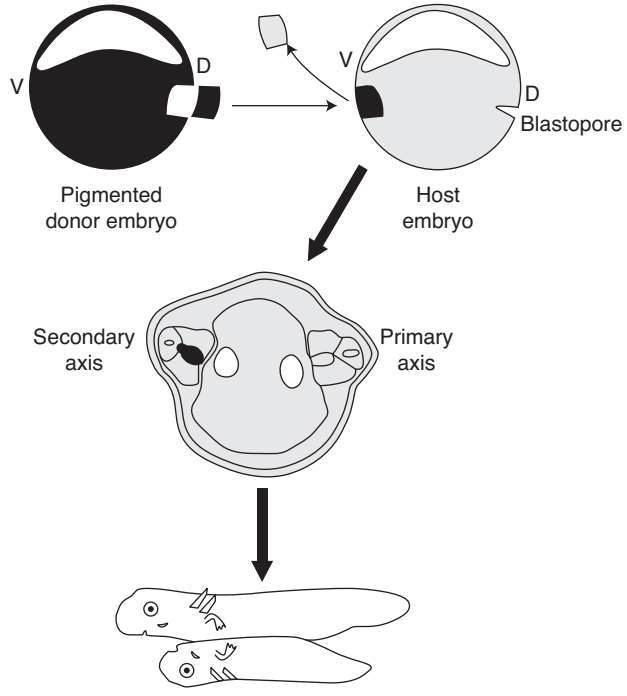


FIGURE 13.2 Organizer transplant experiment of Mangold and Spemann. Tissue around the DBL was removed from one embryo (black) and placed into the ventral side of another (light gray). The transplanted DBL, if large enough, will cause a complete second dorsal axis to form on the host embryo, resulting in twinning. Cross section through the tadpoles shows that the second dorsal axis contains a complete nervous system. Importantly, by using differently pigmented embryos, one can show that the majority of the nervous system in this new dorsal axis is not derived from the transplanted tissue, but rather from host tissue, fated to give rise to ventral tissues in the absence of a graft.

then tissues that are topologically and functionally equivalent to the organizer have been identified in fish (where this region is also called the *shield*), in chick, mouse and rabbit (where it is referred to as the *node*) and, more recently, even in some invertebrates.

Importantly, the secondary axis induced in Spemann and Mangold's experiment contained a complete nervous system that was accurately patterned along its anteroposterior (AP) and DV axes and consisted predominantly of host tissue, indicating that the organizer has both neural-inducing and neural-patterning activity. This observation prompted a quest for the molecules mediating these inductions, in particular for the neural inducer. However, subsequent findings that dead organizer tissue (killed by boiling, freezing, or chemical fixation) as well as many heterologous substances such as methylene blue or even grains of sand were able to induce neural tissue in newt ectoderm led to disillusionment among embryologists, and by the 1960s Spemann's organizer concept had lost much of its impetus. It took another 30 years, the emergence of molecular cloning, and the introduction of a different amphibian model system, for there to be a renaissance of the organizer and an understanding of its mechanism of action.

The Molecular Nature of Organizer Factors and the Default Model for Neural Induction

From the 1950s onward, the African frog *Xenopus laevis* became a popular model organism for biological research.

The ectoderm of the Anuran *Xenopus* turned out to be less easily neuralized than that of the Urodeles used by Spemann and others. A piece of ectoderm isolated from a *Xenopus* embryo before the onset of gastrulation (a so-called "animal cap" explant) will differentiate into epidermis when cultured in the absence of any external factors. In the 1980s, embryologists started to apply the newly developed toolkit of molecular biology to questions of embryonic patterning, and, for the first time, genes that are expressed in specific tissues of an embryo could be identified and visualized by *in situ* hybridization. Rapidly, an increasing number of factors specifically expressed in the organizer were found, and several of those acted as neural inducers when overexpressed in frog embryos by mRNA injection, but how they functioned molecularly remained to be established.

The first step toward an understanding of the molecular mechanism of neural induction came from experiments using a truncated form of an Activin receptor that functioned as a dominant-negative because it lacked the intracellular domain essential for signal transduction. Activin, a member of the transforming growth factor β (TGFβ) superfamily, had previously attracted attention as an inducer of mesoderm in both animal cap explants and entire embryos. Expression of the truncated receptor resulted in neural induction in animal cap explants that would normally form epidermis. This was the first indication that the *inhibition* of a signaling pathway, rather than its activation, could underlie neural induction. It also led to the re-evaluation of an experiment performed by Grunz and Tacke (1989) a few years before, where neural tissue was induced in animal caps that had been dissociated into single cells and reaggregated several hours later: the dissociation of ectodermal cells may have diluted an antineuralizing factor that is normally present in intact ectoderm.

Finally it was demonstrated that bone morphogenetic proteins (BMPs), also members of the TGFβ superfamily, were likely candidates to mediate this antineuralizing signal: (1) they are widely expressed in the embryo before and throughout gastrulation (but their expression becomes downregulated in the neural plate); (2) addition of BMP4 protein to dissociated animal caps prevents neuralization and promotes epidermis formation; (3) overexpression of BMPs by mRNA microinjection results in embryos with a reduced neural plate and expanded epidermis; and (4) the dominant-negative Activin receptor that induced neural induction in animal caps also blocks BMP signaling. The last piece in the puzzle was added when several of the neural-inducing factors secreted by the organizer were found to be inhibitors of BMP signaling: Noggin, Chordin, Follistatin, and Cerberus all bind to BMPs with high affinity, thereby preventing them from activating their receptors, whereas *Xenopus* Nodal-related 3 (Xnr3) inhibits BMP signaling by an as yet unknown mechanism. Taken together, these observations led to the idea that Spemann's organizer induces neural tissue in the dorsal ectoderm by inhibiting BMPs, whereas ventral ectoderm develops into epidermis as it remains exposed to BMPs. An important aspect of this *default model* for neural induction (Muñoz-Sanjuán & Brivanlou, 2002) is that ectoderm will inherently (by default) form neural tissue unless it is exposed to antineuralizing signals—a notion that could explain why so many "neural inducers" were found in the period following Spemann and Mangold's discovery (Fig. 13.3).

The interplay between BMPs and their inhibitors is not limited to the ectodermal germ layer. BMPs were also found to ventralize, whereas organizer-derived BMP inhibitors dorsalize the mesoderm. Thus, a main function of the organizer is to create a sink for BMPs that normally establish ventral identity. Surprisingly, this molecular mechanism appears to be conserved far beyond the deuterostome superphylum, as the formation of the ventral nerve cord in the fruit fly *Drosophila melanogaster* depends on the inhibition of Decapentaplegic (a fly BMP) by the Chordin ortholog Short gastrulation (Bier, 1997).

Due to the extensive redundancy of BMP inhibitors secreted by the organizer, the genetic evidence for their requirement is comparably weak. For example, a loss-of-function mutation in the zebrafish ortholog of *chordin* causes only a moderate reduction in the size of the neural plate (Schulte-Merker, Lee, McMahon, & Hammerschmidt, 1997). Similarly, mice lacking either *Chordin* or *Noggin* function (see Box 13.1) display a relatively normal CNS. The simultaneous deletion of two BMP inhibitors in *Chordin/Noggin* double knockout mice leads to severe head and forebrain defects, but neural tissue is still induced in these embryos (Bachiller et al., 2000). Only the depletion of three BMP inhibitors—Chordin, Noggin, and Follistatin—by injection of morpholino antisense oligonucleotides directed against the mRNAs of these genes into *Xenopus* embryos resulted in a dramatic loss of neural tissue and dorsal mesoderm, indicating a strict requirement for organizer-derived BMP inhibitors (Khokha, Yeh, Grammer, & Harland, 2005). Conversely, the simultaneous depletion of four BMPs using the same experimental strategy leads to a dramatic expansion of neural at the expense of epidermal tissue in *Xenopus* (Reversade & De Robertis, 2005).

The default model for neural induction has also received support from work in embryonic stem (ES) cells. There are different ways of converting ES cells into neurons, but the depletion of growth factors is one of these. Furthermore, addition of BMP protein to ES cell cultures is known to block neural and promote epidermal differentiation, highly reminiscent of its role in vertebrate embryos. Recently, a zinc finger transcription factor, Zfp521, was shown to be strictly required for neural differentiation of ES cells, where it functions by recruiting

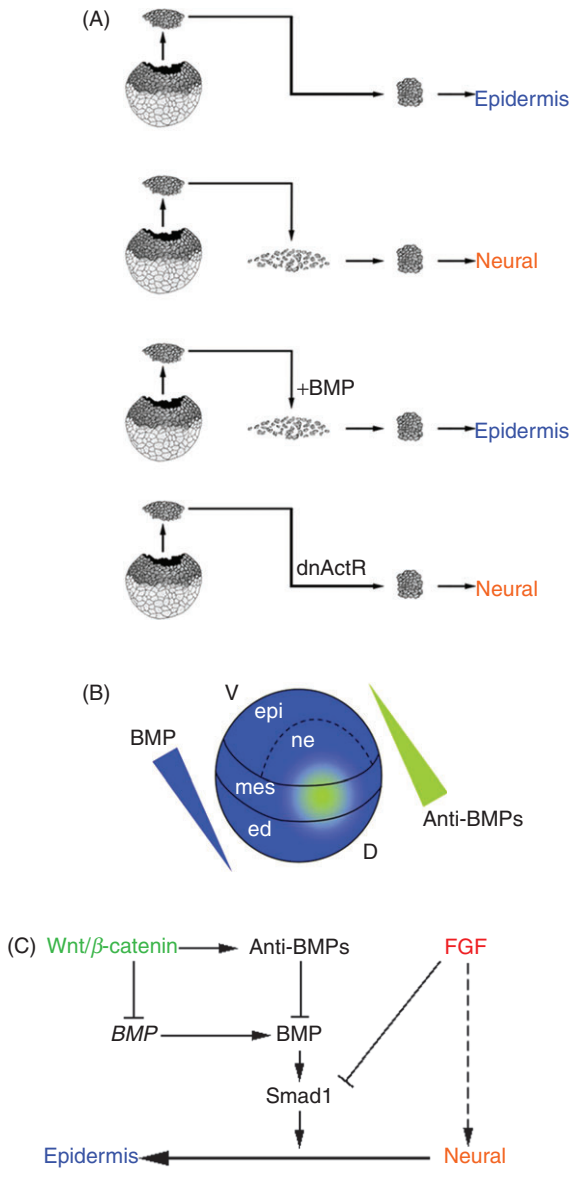


FIGURE 13.3 The BMP signaling pathway and its role in DV patterning of the ectoderm. (A) Experiments in *Xenopus* embryos that led to the default model: culture of animal cap explant results in epidermis differentiation; dissociation for several hours followed by reaggregation of animal cap tissue results in neural induction; the presence of BMPs during dissociation prevents neural induction and promotes epidermis formation; the expression of a dominant-negative Activin receptor results in neural induction even without dissociation. (B) Schematic model for DV patterning of the amphibian embryo. Spemann's organizer (green) on the dorsal (D) side of the embryo produces BMP antagonists that counteract the activity of BMPs (blue) on the ventral side (V). This results in a BMP activity gradient that patterns the DV axis of the embryo. Ventral ectoderm develops into epidermis (epi) under the influence of BMPs, whereas the neural plate (neurectoderm, ne) is induced on the dorsal side through BMP inhibition; mes, mesoderm; ed, endoderm. (C) The default model for neural induction including recent modifications: BMPs induce epidermal fate and inhibit neural induction via Smad1 activity; BMP inhibitors act as neural inducers by blocking BMPs; FGFs act as neural inducers by counteracting Smad1 and via BMP-independent mechanisms; Wnt/β-catenin signaling predisposes ectoderm for neural induction by both preventing the transcription of *Bmp* genes and inducing the expression of BMP inhibitors.

the transcriptional co-activator p300 to the promoters of multiple early neural genes. Forced expression of Zfp521 neuralizes ES cells even in the presence of BMPs. This discovery highlights the importance of intracellular factors in executing the effects of extracellular signals that drive cell fate changes, such as neural induction (Kamiya et al., 2011).

Challenges to the Default Model and Additional Pathways Involved in Neural Induction

Although the fundamentals of the default model have been widely accepted, several recent studies have added complexity to our current view of neural induction. For example, the analysis of fish, frog, and mouse embryos in which Nodal signaling is impaired has cast doubt on the importance of Spemann's organizer in neural induction. Nodal, a TGFβ molecule similar to Activin, is essential for mesoderm and endoderm induction and thus for organizer formation. Surprisingly, Nodal-deficient embryos still form a relatively well-patterned CNS, suggesting that the organizer is dispensable for neural induction. One possible explanation for this phenotype could be that these embryos not only lack an organizer but also most nonorganizer mesoderm and endoderm, which is a likely source of antineuralizing signals. However, BMPs that are expressed in the ectoderm should be sufficient to prevent neural induction in embryos that lack Nodal. Work in frog embryos demonstrated that the Wnt/β-catenin signaling pathway that is known to establish dorsal identity at very early stages of development can induce neural tissue by (1) repressing the transcription of *bmp4* and (2) inducing the expression of *chordin*, *noggin*, *follistatin*, *Xnr3*, and *cerberus* in the dorsal part of the embryo, even in the absence of an organizer. This activity of Wnt/β-catenin signaling occurs before the onset of gastrulation, possibly priming the dorsal ectoderm for neural induction by later signals from the organizer (Baker, Beddington, & Harland, 1999; Wessely, Agius, Oelgeschläger, Pera, & De Robertis, 2001).

In chick embryos, misexpression of *Chordin* fails to induce ectopic neural tissue in regions outside of the presumptive neural plate, whereas misexpression of *Bmps* fails to block neural induction, calling the role of BMPs and their inhibitors into question. Fibroblast growth factors (FGFs) are expressed in the precursors of the chick organizer and have been shown to be both required and sufficient to induce early neural marker genes, leading to a model where FGF signaling initiates neural induction before the onset of gastrulation by inducing a labile early neural character and the BMP inhibitors secreted by the organizer stabilize neural fate in such "primed" ectodermal cells at later stages. How FGFs induce neural tissue remains somewhat controversial.

BOX 13.1

TRANSGENIC MICE AND ENGINEERED MUTATIONS

Rapid and complementary advances in the fields of molecular biology and experimental embryology have combined to offer neuroscience researchers unprecedented power to manipulate the mammalian genome. The technologies used for these manipulations have been worked out primarily in the laboratory mouse and fall into two basic classes: those used for transgenic mice and those used for embryonic stem cell chimeric mice.

Transgenic mice are created by the injection of a cloned DNA fragment into the male pronucleus of a recently fertilized mouse embryo. The fragment will integrate into the host genome and be passed through subsequent mitoses to all the cells of the adult, including the gametes. The integrated DNA fragment, now known as a "transgene," usually is engineered to contain a promoter and associated regulatory sequences, a structural gene, and a 3′ polyadenylation signal. Transgenes add to the genome. As a genetic element, the chromosomal site of integration is random, and there is no wild-type allele on the sister chromosome. As an expressed locus, the transgene message is made over and above the endogenous gene expression pattern. This technique can be used as both an analytical and an experimental tool. Used as an analytical tool, the potency of a certain genetic element to direct cell- or tissue-specific gene expression can be determined by using the element to regulate marker genes such as β-galactosidase or green fluorescent protein. The genetic elements that regulate the temporal and spatial expression pattern of *Hox* genes in hindbrain, tyrosine hydroxylase in adrenergic neurons, and L7 in Purkinje cells, have all been explored by this means. Used as an experimental tool, transgenes can exploit a genetic element with known specificity to deliver a gene product to an ectopic cell site or developmental time. Thus, the PDGF promoter has been used to drive the expression of human β-amyloid precursor protein, the L7 promoter has been used to deliver diphtheria toxin to differentiating Purkinje cells, and the β-actin promoter has been used to deliver *Hox-A1* to inappropriate sites in the developing embryo.

Embryonic stem cells (ES cells) are stable cell lines derived from the inner cell mass of the preimplantation embryo. They are totipotent, which means that if they are introduced into a host embryo, they can contribute to all cell types in the resulting chimera (including gametes). The use of homologous recombination in ES cells allows changes to be engineered in specific genetic loci in culture. By using modified cells to create chimeras, changes can be introduced into the mouse germline and propagated as new mutations. The mutations can be insertions, deletions, modifications, or any combination of the three. When the engineered insertion/deletion disables the normal allele, the resulting mutation often is referred to by the slang term "knockout." These techniques alter the genome. As a genetic element, the engineered locus replaces a specific gene locus, and there is a normal wild-type allele on the sister chromosome. Used in this way, a knockout mutation can be used to model an inherited disease. Lesch-Nyhan (HPRT-null), ataxia-telangectasia (ATM-null), and fragile-X mental retardation syndrome (FMR1-null) have all been modeled in this way. As an expressed locus, the knockout transcript is made instead of the wild-type gene product. This technique can thus be used to create a modified locus such that the targeted gene is mutated rather than destroyed, in the same way an endogenous transcript of one gene can be replaced with that of a different one. Thus, sequences encoding *Engrailed-2* have been inserted into the *Engrailed-1* locus in such a way that the *Engrailed-1* transcript is lost (a null mutation) and *Engrailed-2* is made in its place.

These are only some of the ways in which the powerful new technologies of transgenic and knockout mice are providing powerful genetic tools for use in neuroscience research.

In chick, FGF signaling induces the intracellular protein Churchill that activates Smad-interacting protein 1 (Sip1). Sip1 inhibits Smad proteins, the intracellular transducers of TGFβ signals, and could therefore block both mesoderm-inducing Nodal signaling and antineuralizing BMP signaling (Stern, 2006).

Smad1, the main transducer of the BMP signaling pathway, has phosphorylation sites for mitogen-activated protein kinase (MAPK), which is activated by FGFs as well as some other neural inducing factors such as insulin-like growth factor (IGF) and hepatocyte growth factor (HGF). Phosphorylation of Smad1 by MAPK inactivates its signaling function, providing a further explanation for the neural-inducing activity of FGFs. However, some effects of FGFs, such as the induction of early markers of neural progenitors of the *Sox* gene family, are thought to occur independently from BMP signaling.

The De Robertis laboratory has recently discovered that Smad1 is phosphorylated by glycogen synthase kinase 3 (GSK3), which targets it for proteasomal degradation. Since Wnt signaling inhibits GSK3 activity, Wnts may inhibit neural induction by stabilizing Smad1. Indeed, an antineuralizing and epidermis-inducing effect of Wnts has been reported in some chick and frog studies. At first

sight, this seems to be at odds with the idea that Wnt signaling predisposes dorsal ectoderm for neural induction before gastrulation; however, the timing of the Wnt signal likely plays a key role in these differential effects. Intriguingly, MAPK phosphorylation primes Smad1 for phosphorylation by GSK3, suggesting that the neural-inducing MAPK signal initiates its own termination. This study provides a beautiful example as to how a single molecule can integrate three signaling pathways—BMP, FGF, and Wnt—to generate a temporally controlled output (De Robertis, 2006; Fuentealba et al., 2007).

Summary

Classical transplantation experiments showed that Spemann's organizer is able to induce an ectopic neural plate when grafted to the ventral side of a host embryo. Molecular studies have revealed that BMPs block neural induction and that the organizer expresses various BMP inhibitors that function as neural inducers, leading to the default model for neural induction, which suggests that ectodermal cells will become neural as long as they are not exposed to antineuralizing signals. However, complexity was added to this model with the finding that FGF and Wnt signaling establish an early neural potential before the onset of gastrulation, highlighting the necessity to view neural induction as a multistep process. The characterization of transcription factors such as Smad1 and Zfp521 has started to draw attention to the importance of intracellular factors in integrating extracellular signals and in coordinating the temporal progression of neural induction.

EARLY NEURAL PATTERNING

The first step in establishing the morphological and cellular complexity of the mature CNS is the establishment of polarity along the anterior-posterior (AP) and dorsal-ventral (DV) axes of the neural plate. At later stages, the neural plate is parcellated into subdivisions on both of these principal axes to give rise to a multitude of different cell types in a spatially organized manner; these then become wired up in a stereotyped fashion and subserve distinct tasks in functional neural networks.

Classical Models for Early Neural Patterning—Establishment of AP and DV Axes

The ectopic CNS induced in Spemann's organizer experiment was patterned along its AP and DV axes, indicating that the organizer is not only able to induce neural fate, but also to mediate neural patterning. Spemann noted that inductions by the organizer differ depending on the developmental stage of the donor embryo. When he transplanted the DBL from an embryo at early gastrula stage, the graft induced a complete secondary embryonic axis; however, when he transplanted the DBL from a late gastrula donor, the graft only induced ectopic trunk and tail structures (Fig. 13.4A). A possible explanation for this

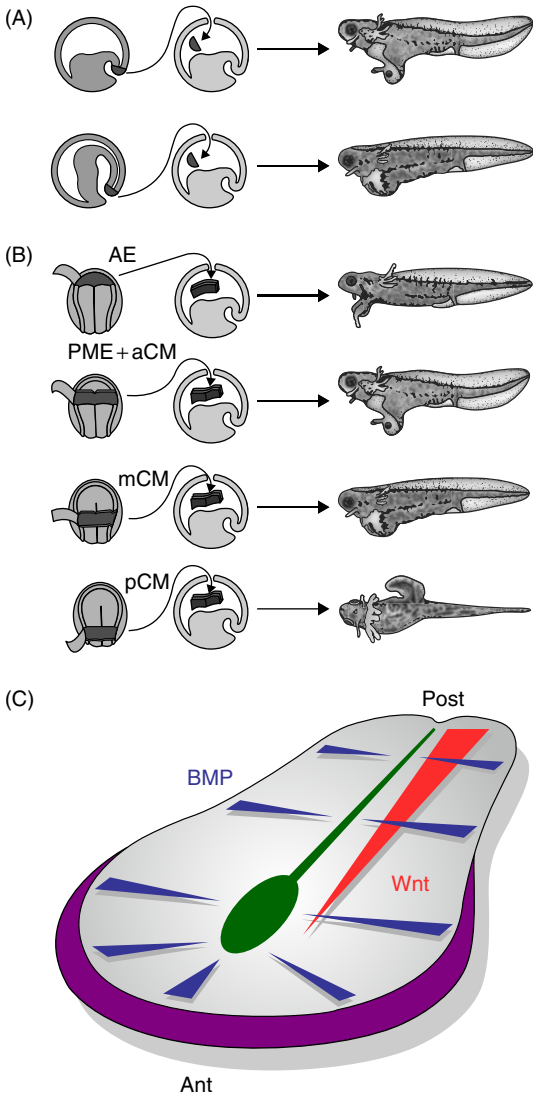


FIGURE 13.4 Models for early AP patterning of the CNS. (A) A graft of the DBL at early gastrula stages (EGO) results in the induction of a complete secondary axis, whereas a late gastrula DBL graft (LGO) only results in trunk and tail induction. (B) Anteriormost sections of the AME that contain mostly AE induce only very anterior ectopic structures; sections containing PME and anterior CM induce heads, whereas more posterior sections induce trunks and tails, leading to the idea of regionally specific inductions. (C) Model for pattern formation in the early neural plate: graded Wnt signaling (red) specifies the anteroposterior axis, whereas graded BMP signaling (blue) specifies the mediolateral axis of the neural plate. These gradients are shaped by the expression of BMP inhibitors in the AME (green) and of Wnt inhibitors in the anterior AME and in the ANB (purple).

observation is that the early gastrula organizer (EGO) contains a population of cells that release a head-inducing signal and that this subpopulation of organizer cells is lost from the late gastrula organizer (LGO).

During gastrulation, cells in the organizer region involute through the DBL, move toward the future anterior end of the embryo, and stretch out to form the axial mesendoderm (AME), which, from anterior to posterior, consists of the anterior endoderm (AE), prechordal mesendoderm (PME), and chordamesoderm (CM). Later on, the AE gives rise to the foregut, thyroid gland, and liver; the PME gives rise to the prechordal plate that underlies the presumptive forebrain region; and the CM gives rise to the notochord, a rod-shaped structure that acts as a signaling center involved in neural tube polarization (see following). The parts of the neural plate/tube that are found above the PME and CM are also referred to as prechordal and epichordal CNS, respectively.

Hilde Mangold's husband Otto dissected pieces of dorsal mesoderm that contained different AP segments of the AME from an early neurula stage embryo and tested their inductive potential by grafting them to gastrula stage host embryos. The anteriormost pieces (containing AE) only induced very anterior structures such as balancers and parts of the oral apparatus; tissue containing PME or anterior CM frequently induced ectopic head structures, including forebrain, eyes, otic vesicles, and midbrain; and tissue containing posterior CM predominantly induced trunk and tail structures, including spinal cord (Fig. 13.4B). Taken together, Spemann's and O. Mangold's experiments suggested that the EGO can induce a full embryonic axis because it contains progenitors of AE, PME, and CM, whereas the LGO can only induce trunk and tail structures because AE and PME progenitors have left the organizer region by that stage. Spemann proposed that inductions by the organizer are *regionally specific* due to the existence of functionally separable head, trunk, and tail organizers. The recent analysis of gene expression in the organizer supports the idea that AE, PME, and CM precursors are unevenly distributed within the EGO—even before a visible DBL has formed (Zoltewicz & Gerhart, 1997).

In the 1950s, Spemann's model had lost its popularity, and embryologists searched for alternative explanations for how the emerging CNS is patterned. Nieuwkoop and his colleagues in Utrecht implanted folds of ectodermal tissue at different AP levels into the prospective neural plate of *Xenopus* embryos. They found that folds implanted at the forebrain level tend to differentiate exclusively into anterior neural structures, whereas folds implanted more posteriorly tend to differentiate into posterior neural tissue at their base, close to the host's neural plate, and into anterior neural tissue more distally. At the time, the idea that tissues could be patterned by morphogens, diffusible factors that induce multiple different cell fates as a function of their dose, had become very popular. Nieuwkoop proposed that a planar signal that travels within the neural plate induces posterior identity in a dose-dependent, morphogen-like fashion. This posteriorizing factor—which was called *the transformer* by Nieuwkoop—is absent from the anterior end of the neural plate where the forebrain forms and gradually increases posteriorly, specifying midbrain, hindbrain, and spinal cord, depending on its dose.

Molecular Basis of Early AP Patterning

The first candidate for Nieuwkoop's transformer was retinoic acid (RA). The teratogenic activity of RA had been known for a long time, but when molecular markers for different regions of the nervous system became available in the 1980s, these effects were tested in a quantitative fashion. Treatment of frog embryos with increasing doses of RA resulted in a progressive loss of head structures and a concomitant upregulation of posterior neural marker genes at the expense of anterior markers. As a small, diffusible molecule that can easily cross cell membranes, RA was an attractive candidate to posteriorize the neural plate in a morphogen-like fashion. RA is synthesized by enzymes of the retinaldehyde dehydrogenase family, and only one member of this family, Raldh2, is expressed in early embryos. In *Raldh2* knockout mice as well as *raldh2* mutant zebrafish, markers of early AP neural regionalization are expressed normally, although patterning defects in the hindbrain become evident at later stages of neural development (see following). Thus, RA is unlikely to be the endogenous transformer, and its neural patterning function appears to be limited to the hindbrain (Maden, 2007).

In addition to their early role in neural induction, FGFs have been proposed also to posteriorize neural tissue. Overexpression of FGFs results in headless frog and fish embryos and treatment of neuralized animal caps with increasing doses of FGF protein results in a progressive induction of posterior neural markers. No genetic loss-of-function in mouse or zebrafish embryos has supported a requirement for FGFs in neural posteriorization, but this could be due to the severe early defects that various FGF mutants display. However, expression of a truncated, dominant-negative FGF receptor in frog embryos results in dramatic posterior truncations, indicating that FGFs are indeed required for posterior development. Yet, these embryos lack not only the posterior CNS but also all other posterior tissues, raising the question as to whether the effects on posterior neural development are direct or a secondary consequence of a lack of a posteriorizing factor secreted from, for example, the posterior mesoderm. Nevertheless, MAPK activity can be detected in the posterior neural plate of gastrulating *Xenopus* embryos using an antibody that specifically recognizes the activated form of this FGF transducer (Mason, 2007).

In higher vertebrates such as mice and chicks the posterior neuraxis is only formed late during gastrulation and is progressively laid down as the node regresses. Thus, AP patterning of the CNS is more protracted in these embryos than in frog and fish and requires a more complex spatiotemporal coordination. Experiments in chick embryos showed that FGF8, which is expressed in the regressing node, maintains a pool of stem cells that is required for posterior axis extension (Diez del Corral & Storey, 2004; Mathis, Kulesa, & Fraser, 2001). Thus, it is possible that the role of FGF signaling is to determine when cells leave the node and that those cells that remain in the stem zone acquire more posterior identities as they are exposed to another posteriorizing factor, the transformer, for longer.

Signaling factors of the Wnt family are currently the best candidates to mediate neural posteriorization in a manner consistent with Nieuwkoop's transformer. These secreted glycoproteins can activate a variety of signaling pathways of which the Wnt/β-catenin pathway is arguably best understood. Overexpression of Wnts antagonizes anterior development and promotes posterior development in frog, fish, chick, and mouse embryos. Treatment of neuralized animal caps with different doses of Wnts results in a stepwise posteriorization and implanting Wnt-expressing cells into the anterior neural plate of gastrulating *Xenopus* embryos results in the outgrowth of a mini-neuraxis around the implanted cells, indicating that they act as an organizer of posterior development. Finally, a graded activation of the Wnt signaling pathway can be detected along the AP axis of the neural plate by quantifying the nuclear localization of β-catenin, a hallmark of Wnt activity (Kiecker & Niehrs, 2001).

Several factors expressed by Spemann's organizer were found to be Wnt inhibitors, and these remain expressed in the anterior but are absent from the posterior AME during gastrulation. Cerberus is a multifunctional inhibitor that sequesters BMPs, Nodal, and Wnts and is expressed in the AE; Dickkopf1 (Dkk1) blocks the Wnt/β-catenin pathway at the level of receptor activation and is expressed in the PME; and Frzb1 sequesters Wnts by direct binding and is expressed in the PME and in the anterior CM. The strongest evidence that Wnt inhibition from the organizer is required for anterior neural development comes from the genetic knockout of *Dkk1* in mice that results in embryos lacking both forebrain and midbrain. A comparable phenotype is observed in frog embryos following injection of a neutralizing antibody directed against Dkk1 (Niehrs, 2004).

Thus, tissues attributed with head organizer activity by Spemann and Mangold (EGO, PME, and anterior CM) express Wnt antagonists, whereas trunk/tail organizers (LGO, posterior CM) do not. This leads us to a reconciliation of Spemann and Mangold's model of regionally specific inductions and Nieuwkoop's gradient-based model: the head organizer expresses Wnt inhibitors that help to establish a posteriorizing gradient of Wnt activity along the AP axis of the neural plate; anterior neural tissue is induced in the prechordal neural plate where levels of Wnt signaling are lowest. Just as an electron can act both as a wave and as a particle, AP patterning is regulated by organizers or gradients, depending on the experimental approach.

Other Tissues Involved in AP Patterning

An extraembryonic tissue, the anterior visceral endoderm (AVE), underlies the presumptive anterior neural plate in mouse embryos and becomes displaced anteriorly by the extending AME during gastrulation. The AVE expresses a range of factors that are also expressed in the organizer, suggesting that it could play a role in neural induction and/or patterning. In fact, when initial attempts in grafting the mouse EGO or LGO failed to induce heads it was suggested that head and tail organizer activity might be split between the AVE and the node in mouse embryos. However, later studies demonstrated that the mouse node is able to induce a complete secondary axis upon transplantation at mid-gastrula stage. The mouse AVE and its chick equivalent, the anterior hypoblast, are not sufficient to stably induce anterior neural markers when grafted to ectopic locations; yet the AVE can suppress posterior neural marker genes and the anterior hypoblast transiently induces early neural markers, suggesting that they may act permissively to prime the anterior neural plate for later signals from the organizer that stabilize the early labile anterior neural fate. Notably, the AVE expresses the BMP/Nodal/Wnt inhibitor Cerberus and the Wnt inhibitor Dkk1 that are likely to mediate the function of the AVE.

A single row of cells at the anterior border of the neural plate (ANB) of the zebrafish gastrula embryo is also involved in establishing anterior neural identity, as its surgical removal results in a loss of anterior character and, conversely, transplantation of ANB cells to more posterior locations leads to an ectopic induction of anterior neural fates (Houart, Westerfield, & Wilson, 1998). The ANB was found to express a Wnt inhibitor related to Frzb1, Tlc, whose activity mimics ANB function. Thus, tissues other than the anterior AME can exert head organizer function—but their unifying feature is the expression of Wnt inhibitors that counteract the posteriorizing influence of Wnts.

Early DV Patterning

Whereas Wnt inhibitors are only expressed in the anterior AME, the BMP inhibitors secreted by the EGO remain expressed along the entire AP extent of the AME. Since the AME underlies the neural plate along its midline, it can act

as a sink for BMPs expressed in the ectoderm surrounding the neural plate. Experiments in zebrafish have suggested that a low/medial-high/lateral gradient of BMP activity is established across the neural plate that specifies different mediolateral neural fates in a morphogen-like fashion depending on the BMP dose (Barth et al., 1999).

Summary

Classical models for early AP neural patterning suggested the presence of qualitatively distinct head, trunk, and tail organizers or the quantitative activity of a posteriorizing morphogen. Both models are now reconciled by the emerging molecular mechanism of this process: Wnts posteriorize the neural plate in a dose-dependent fashion, and tissues with head organizer activity protect the anterior neural plate from this posteriorizing influence by secreting Wnt inhibitors, thereby generating an AP gradient of Wnt activity across the neural plate. FGFs and RA also contribute to neural posteriorization. A gradient of BMP activity is established by the interplay between BMPs from the ectoderm and BMP inhibitors from the AME that patterns the mediolateral axis of the neural plate. Thus, quasi-orthogonal BMP and Wnt gradients establish a near Cartesian coordinate system within the neural plate and neural cells become allocated to a specific fate according to their position within this grid (Fig. 13.4C). However, it is clear that a mechanism solely based on global gradients can only establish a crude initial pattern. We will see later that additional mechanisms are required to refine this initial pattern.

Early Neural Morphogenesis

Striking morphological changes in cellular and tissue structure accompany the processes of neural induction and neurulation. At the cellular level, ectoderm cells lose their cuboidal shape and elongate into the pseudostratified morphology that is characteristic of neural precursor cells along the entire axis of the embryo. At a tissue level, the neuroepithelium of the neural plate undergoes the complex morphogenetic movements of neurulation to form the neural tube, which invaginates, dives into the embryo, pinches off from the surrounding ectoderm, and forms a separate tissue. Due to the geometric changes underlying this process, the initial mediolateral polarity of the neural plate becomes translated into the ventrodorsal polarity of the neural tube. Neurulation is an exceedingly complex process and prone to failure for multiple reasons. Indeed, a failure to close the neural tube has devastating downstream consequences on neural development and is the cause of neural tube defects (NTDs), one of the most common classes of human birth defects, occurring in approximately one in 1000 live births.

In higher vertebrates, neurulation can be divided into two phases that differ in their morphogenetic movements: a primary phase involving the brain and most of the spinal cord and a secondary phase involving more posterior regions of the spinal cord. While we are far from understanding the tissue mechanics that underlie neurulation, some progress has been made in characterizing specific cellular behaviors within subregions of the neuroepithelium. Among the earliest of these are the convergent-extension (CE) movements during gastrulation and neurulation that drive elongation of the embryo along its AP axis (Fig. 13.5A). CE is driven by organized cell intercalation much like the behavior of movie-goers when they exit a theater using only a central aisle (Keller, 2002). In a similar manner, neural plate cells intercalate along the mediolateral axis of the neural plate, resulting in a narrowing of the ML axis and simultaneous extension of the AP axis of the embryo.

CE is known to require planar cell polarity (PCP) signaling, a pathway originally identified in *Drosophila* where PCP mutations cause defects in the planar organization of bristles on the wing (Wang & Nathans, 2007). Proteins of the PCP pathway localize asymmetrically to opposite

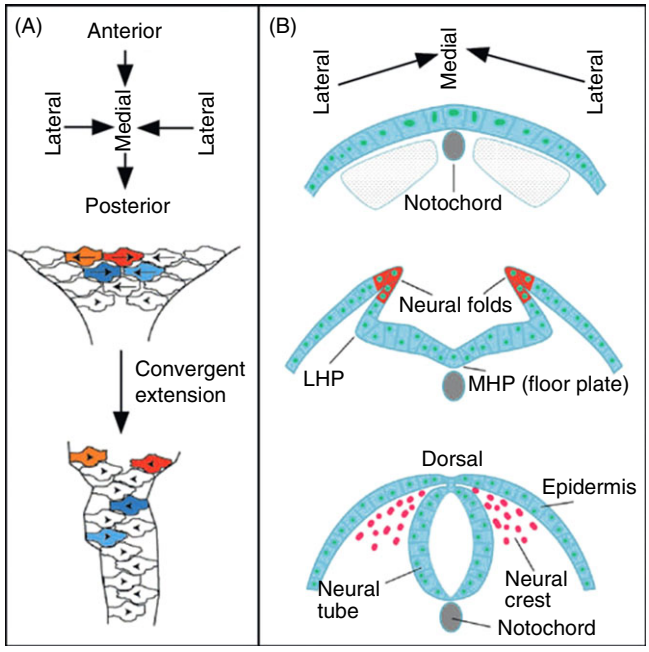


FIGURE 13.5 Morphogenesis of the neural plate and tube. (A) The formation of the neural plate and tube requires an organized intercalation of cells along the medial-lateral axis (convergence), which extends the neural anlage along the anterior-posterior axis (extension). The polarized cell behaviors that drive this process require the PCP pathway. (B) Neurulation of the neural plate into a tube requires several hinge points along the DV axis, including one at the forming ventral midline (MHP) which becomes the floor plate, as well as more dorsal ones (LHP). These hinge points allow the neural plate to buckle and the neural folds to elevate and fuse at the dorsal midline. At the same time, the neural crest cells form (red) and emigrate.

sides of cells within the plane of an epithelium, where they influence the planar orientation of the cytoskeleton. Vertebrate orthologs of these proteins are required during CE to organize the cytoskeleton of migrating cells along the mediolateral axis of the neural plate in a manner that orients polarized cell rearrangements. Consequentially, experimental interference with PCP signaling or mutations of PCP pathway genes disable CE, resulting in an abnormally wide neural plate, a subsequent failure to elevate the neural folds at the edges, and an open neural tube.

CE movements are far more pronounced in the posterior neural plate compared to anterior regions, resulting in the distinctive elongation of the hindbrain and spinal cord. Thus, patterning and morphogenesis are intimately linked during CNS development and experimental perturbation of one typically also affects the other. For example, the implantation of Wnt-expressing cells into the anterior neural plate of a frog embryo results in an induction of posterior neural fate, but also in the outgrowth of a mini-neuraxis (see above).

As CE extends the neural plate along the AP axis, other localized cell behaviors along the DV axis contribute to the buckling of the plate into a tube. Hinge points form at the midline of the neural plate as well as at more dorsal positions (Fig. 13.5B) by apical constriction of actin microfilaments in neuropithelial cells that produces a wedge-shaped morphology. Accordingly, mutations in various proteins that regulate microfilament dynamics are commonly associated with NTDs in the mouse. Shaping of the neural tube also requires localized regions of apoptosis, the lack of which can lead to NTDs. Finally, the neural tube closes with the fusion of the neural folds, forming the dorsal midline of the neural tube. The neural folds give rise to the neural crest, a population of migrating neural progenitors. Mutations in the mouse that alter neural crest formation and emigration are also capable of producing NTDs. Even with this short summary one can appreciate that integrating these various aspects of cell behavior into a unifying mechanical model of neurulation is a formidable task—although one that is required for a better understanding of the etiology of several human birth defects.

Summary

The formation of the vertebrate nervous system is accompanied by striking changes in cell and tissue morphology. Neural tube formation involves the complex morphogenetic process of neurulation that requires coordinated changes in cell shape, cell migration, cell death, and cell–cell contacts. One prominent cell behavior during neural plate and tube formation is CE that depends on PCP signaling and narrows the neural anlage mediolaterally and extends it along the AP axis by cell intercalation.

REGIONALIZATION OF THE CENTRAL NERVOUS SYSTEM

Following the establishment of polarity and the delineation of prechordal and epichordal regions of the neural plate, the AP axis becomes further regionalized into smaller domains, as revealed by the expression of developmental control genes. This process of progressive regional refinement involves two general classes of mechanism—the partitioning of the neuroepithelium into small modules or segments in which development can proceed with a degree of autonomy, as seen in the hindbrain, and the establishment of local organizers as sources of diffusible factors (morphogens), as seen at the junction between the midbrain and hindbrain and at specific locations in the forebrain. In both cases, a conspicuous and important feature is the setting up of narrow boundary regions that either partition the segments or provide a specialized signaling function, or both. We will illustrate these patterning mechanisms by reference to selected examples. Our intention is to outline general principles rather than to provide a comprehensive review of nervous system patterning.

Hindbrain Segmentation

Shortly after neural tube closure, a series of seven varicosities appear along the AP axis of the hindbrain region. Although transient, these *rhombomeres* are true segments with defined sets of precursor cells allocated to an axially repeated set of similar modules (Lumsden & Keynes, 1989). Thus, the earliest formed neurons are laid out in stripes that match the morphological repeat pattern, with neurogenesis starting within the confines of alternate, even-numbered rhombomeres, and only appearing later in odd-numbered rhombomeres—a "two-segment repeat" pattern (Fig. 13.6). Later in development, these segmental origins become obscured as the neuronal nuclei condense and migrate bodily to new positions.

Segmentation is a strategy for specifying the pattern of developing structures that has been adopted independently by many animal phyla, offering the advantages that organizational fields remain small and specializations of cell type and pattern can be generated as individual segmental variations on the repetitive theme. In a segmented system, precise boundaries can be set for both cellular assemblies and realms of gene action.

The neural tube cannot become segmented in the same way as the mesoderm, where physically separate somites are formed, because the tube has to retain epithelial continuity, not least as a conduit for extending axons. Rather, the process must involve some mechanism that restricts cell mingling. Vital dye and genetic marking experiments have revealed such a *cell lineage restriction* mechanism in the hindbrain: clonal descendants of marked cells disperse

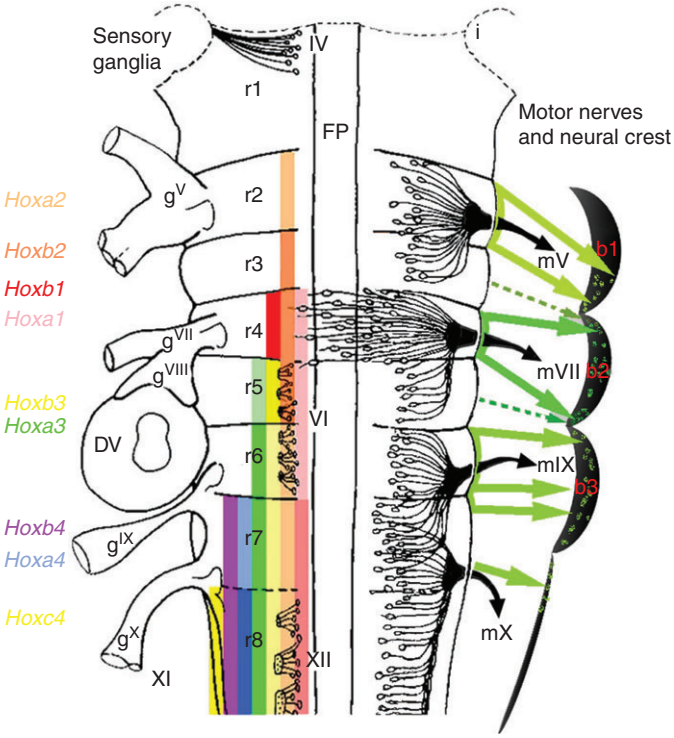


FIGURE 13.6 Distribution of motor neuronal types in the chick embryo hindbrain in relation to rhombomeres. Shown on the right side are branchiomotor neurons, forming in r2 + r3 (Vth cranial nerve, trigeminal), r4 + r5 (VIIth nerve, facial), and r6 + r7 (IXth nerve, glossopharyngeal), and contralaterally migrating efferent neurons of the VIIIth nerve (vestibuloacoustic), which are in the floor plate (FP) of r4 at the stage shown. Shown on the left side are somatic motor neurons, forming in r1 (IVth nerve, trochlear), r5 + r6 (VIth nerve, abducens), and r8 (XIIth nerve, hypoglossal). Cranial nerve entry/exit points and sensory ganglia associated with r2 (trigeminal), r4 (geniculate, vestibuloacoustic), r6 (superior), and r7 (jugular) are shown, as is the otic vesicle (ov). Colored bars represent the AP extent of *Hox* gene expression domains; note that one of these, *Hoxb1*, is expressed at a high level only in r4. *Modified from Lumsden and Keynes (1989).*

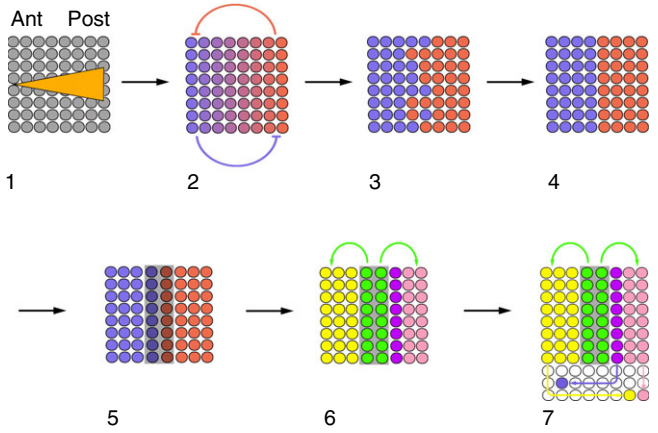


FIGURE 13.7 Model for neural boundary formation. 1. An initially uniform sheet of cells is polarized by a global signaling gradient (yellow)—for example, by a posterior-to-anterior gradient of WNTs. 2. This results in a coarse prepattern of transcription factor expression, with blue induced at low WNT signal level and red at a higher level. 3. Mutual cross-repression between blue and red factors establishes two distinct populations of cells separated at a fuzzy interface. 4. Cell sorting processes, depending on affinity molecules regulated by the red and blue transcription factors, act by minimizing contact between red and blue cells; a sharp, straight interface results. 5. Interaction between red and blue cells at the interface induces a boundary cell phenotype and boundary-specific markers are expressed (shaded area). 6. The boundary cells then express a signal molecule (green) that spreads laterally. Asymmetric responses by receiving cells to the green morphogen may depend either on prepatterned difference in competence (blue cells become yellow, and red cells become either both mauve and pink), and on morphogen concentration (mauve at high level, pink at lower level). The diffusion of a morphogen from a local source also governs the timing and spatial extent of responses. Later (7), postmitotic cells may migrate to new positions in the non-lineage-restricted mantle layer. *From Kiecker and Lumsden (2005).*

widely within the neuroepithelium, but the spreading cell clone always remains within a single rhombomere, confined at its boundaries (Fraser, Keynes, & Lumsden, 1990; Jimenez-Guri et al., 2010). The vertebrate hindbrain thus shares with insects the phenomenon of modular construction using cell-tight developmental compartments, where lineage restriction prevents cells wandering from one rhombomere to another and blurring the resulting pattern.

What mechanism is responsible for segregating hindbrain cells into compartments? One possibility is that cells of adjacent rhombomeres express surface molecules that favor affinity between the cells within a rhombomere but reduce affinity for cells in adjacent rhombomeres. Consistent with this, cultured cell aggregate experiments have shown that cells from even-numbered rhombomeres mix evenly with cells from other evens (as do odds with odds), whereas odd and even cells segregate from each other (Wizenmann & Lumsden, 1997). Thus, rhombomeres may partition according to an adhesion differential, obeying a two-segment repeat rule. Candidate molecules for mediating reduced intercellular affinity or adhesion at rhombomere boundaries include the Eph family of receptor tyrosine kinases and their ephrin ligands (Xu, Mellitzer, Robinson, & Wilkinson, 1999). Eph receptors are expressed in odd-numbered pre-rhombomeres, whereas their ligands are expressed in evens. This mutually exclusive pattern means that ligand–receptor interaction can only occur at forming boundaries, where it results in sharpening of the initially fuzzy interface between the adjacent domains (Fig. 13.7). Little is known about how the segmental pattern of the hindbrain is initially set out. Among the few known candidate regulators of segmental pattern is the zinc finger protein Krox20, an upstream transcriptional regulator of Eph receptors, which is expressed in two stripes in the neural plate that later become rhombomere (r)3 and r5.

Summary

Shortly after closure, the hindbrain neural tube becomes subdivided by transverse boundaries to form rhombomeres, a series of repetitive elements or metameres that have compartment properties. The existence of compartmental organization in the hindbrain suggests that cells become specified in segmental groups during their confinement in the ventricular zone and that they initially share a common identity, or ground state. First a row of similar boxes is formed and then the boxes are assigned individual identities according to their positional value.

Definition of AP Pattern by Differential Homeobox Gene Expression

Central to the illumination of vertebrate CNS pattern formation has been the discovery that developmental control genes related to genes with a known patterning role in the *Drosophila* embryo are expressed in spatially restricted domains of the neural plate and tube. These regulatory genes, many of which encode homeodomain-containing transcriptional control proteins, include the *Hox* genes, which specify positional value along the AP body axis of the fly embryo. In both flies and vertebrates the *Hox* genes have a clustered chromosomal organization in which the relative position of a gene in the cluster reflects its boundary of expression along the AP axis. In vertebrates, the *Hox* genes are expressed in overlapping, or nested, domains along the AP axis of the early embryo, with those at the 3′ ends of the cluster being expressed most anteriorly, in the hindbrain, where there is a precise correspondence between their anterior expression borders and the boundaries between the rhombomeres (Fig. 13.8).

Accumulating evidence supports the view that, by encoding positional value, *Hox* genes control the identity and phenotypic specializations of subregions of the epichordal neural tube in which they are expressed. Loss-of-function mutations of anteriorly expressed *Hox* genes result in malformations that represent a transformation of rhombomere identity. In the *Hoxb1* mutant mouse, for example, r4 (where the gene is normally expressed at a high level) loses its r4-specific character and takes on that of r2, a more anterior domain where the gene is not normally expressed (Studer, Lumsden, Ariza-McNaughton, Bradley, & Krumlauf, 1996). Similarly, overexpression of *Hoxb1* in r2 causes it to adopt phenotypic characters of r4 (Bell, Wingate, & Lumsden, 1999).

Anterior to r2, the brain does not express *Hox* genes but displays spatially restricted expression of other transcriptional control genes whose homeoboxes are divergent from the *Hox* type. These genes also are highly conserved between flies and vertebrates. Two homologues of the *Drosophila* segmentation gene, *engrailed (en)*, are expressed in a broad region either side of the midbrain–hindbrain boundary (MHB), which later forms the optic tectum and the cerebellum (Figs. 13.9 and 13.10). Expression of the *En* genes is graded, being strongest at the MHB and declining both anteriorly and posteriorly. Morphological derivatives of the entire domain of *En* expression are deleted in *En1* knockout mice, showing that *En* function is crucially involved in the morphogenetic specification of the region (Wurst, Auerbach, & Joyner, 1994).

The gap genes *orthodenticle* and *empty spiracles*, which function as homeotic selectors in the specification of particular head segments in *Drosophila*, have vertebrate homologues that are expressed in overlapping domains that encompass the entire rostral extremity of the AP axis with the exception of the ventral forebrain. The two *empty spiracles* homologues, *Emx1* and *Emx2*, are forebrain specific, whereas the two *orthodenticle* homologues, *Otx1* and *Otx2*, have wider expression, encompassing both forebrain and midbrain. The *Otx2* knockout mouse has an extreme phenotype that shows the importance of the gene but betrays little about its function—the entire head rostral to r3 is deleted. *Otx2* is also expressed in the PME that lies beneath the anterior ectoderm, where it is required for the induction of anterior neural structures during gastrulation (Simeone, Puelles, & Acampora, 2002). However, the gastrulation phenotype can be rescued by expressing *Otx1* in the visceral endoderm, while the ectoderm remains functionally null for *Otx2*. This reveals the brain-specific

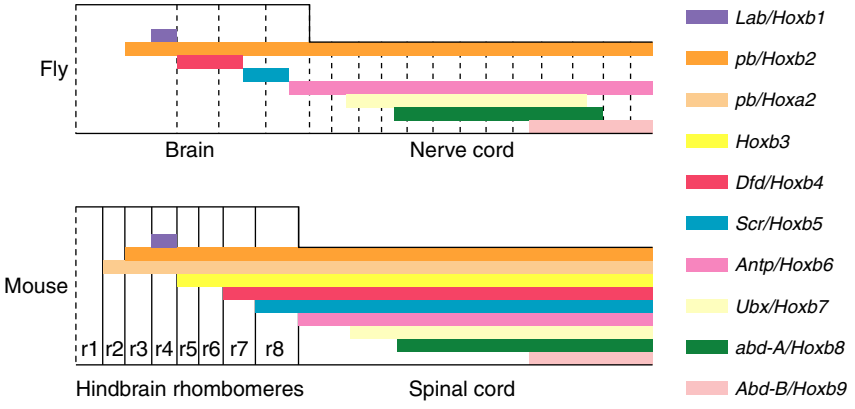


FIGURE 13.8 *Hox* gene expression domains in the CNS of fly and mouse. Nested domains of *Hox* genes along the AP axis of the *Drosophila* CNS closely parallel those of their homologues in mouse. Compare, for example, the fly gene *labial (lab)* with its mouse homologue *Hoxb1*. *Hox* genes specify a positional value along the AP axis, which is interpreted differently in fly and mouse in terms of downstream gene activation, resulting in neural structure; shared between the two organisms is the means of encoding the position of a cell along the AP axis. *From Hirth, Hartmann, and Reichert (1998).*

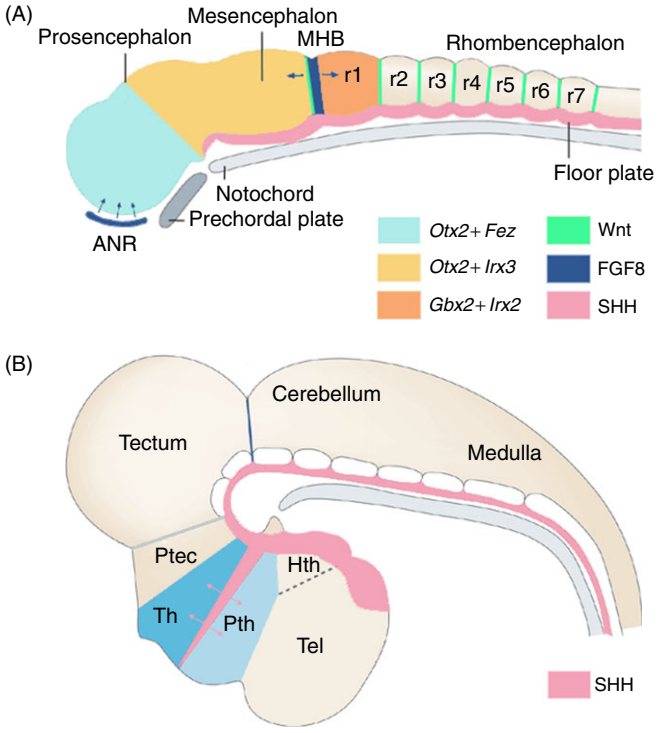


FIGURE 13.9 Boundaries in the avian embryo brain. (A) At an early stage signaling boundaries form at an interface between domains of Otx2 and Gbx2 expression, the midbrain–hindbrain boundary (MHB), and between adjacent rhombomeres (r1–r7). The zona limitans intrathalamica, organizer of the diencephalon, will later form at the interface shown here between Fez and Irx3 transcription factor domains. Dark blue arrows represent FGF8 signaling to both the presumptive optic tectum (mesencephalon) and presumptive cerebellum (r1). The anterior neural ridge (ANR) also secretes FGF, patterning the anterior forebrain. (B) At a later stage, a narrow wedge-shaped boundary region (ZLI) forms in the mid-diencephalon, and expresses Sonic hedgehog (pink), which induces the expression of regional determinants in flanking territories—the thalamus (Th) and prethalamus (Pth). Hth, hypothalamus; Ptec, pretectum; Tel, telencephalon.

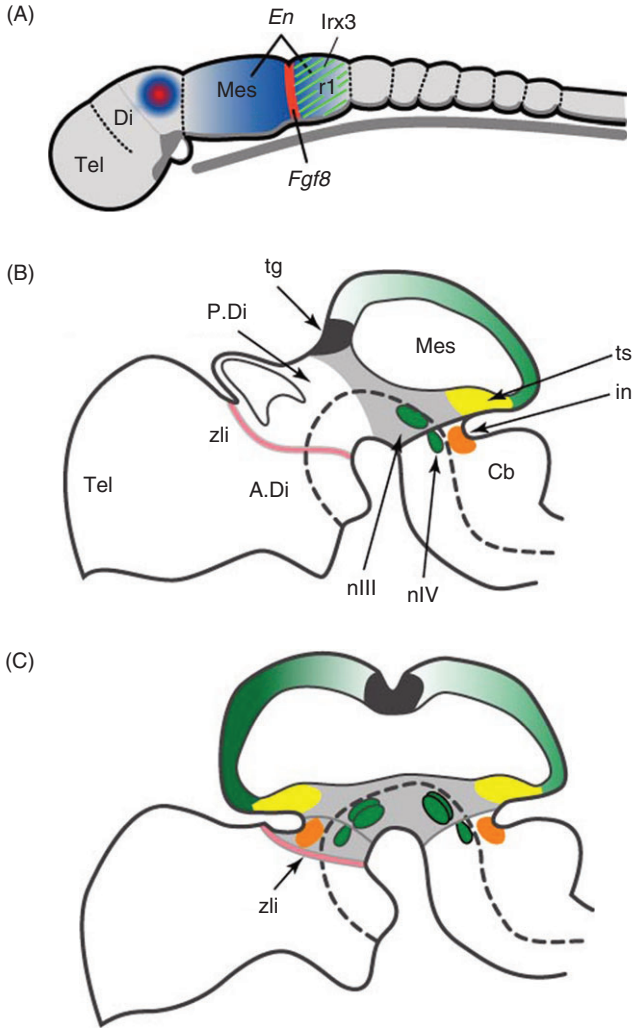


FIGURE 13.10 Role of FGF8 in mid/hindbrain patterning. (A) Implanting a bead that releases FGF8 protein (red) in the posterior diencephalon of an early chick embryo results in the induction of *En* (blue) and the transformation of normal posterior diencephalic territory (B) into midbrain (C). In the treated embryo (C), the posterior diencephalon forms a set of midbrain structures laid out in reverse AP polarity to the normal midbrain. This is thought to be due to the induced anterior-to-posterior gradient of EN protein that is the mirror image of the endogenous EN gradient. Also shown in (A) is the expression domain of the prepattern gene *Irx2* (green hatched), which establishes cerebellar competence in rhombomere 1. A.Di, anterior diencephalon; in, isthmic nuclei; Mes, midbrain; nIII, oculomotor nucleus; nIV, trochlear nucleus; P.Di, posterior diencephalon; r1, rhombomere 1; Tel, telencephalon; tg, tectal griseum; ts, torus semicircularis; zli, zona limitans intrathalamica. *Data from Crossley, Martinez, and Martin (1996) and Matsumoto et al. (2004).*

function of the gene, which is to maintain anterior neural identity; in the absence of neuroectodermal *Otx2*, the anterior brain becomes converted into hindbrain, with an enlarged cerebellum rather than forebrain forming at the anterior end of the CNS.

An important later function of *Otx2* is to set the position of the midbrain–hindbrain boundary (MHB), whose local organizer functions are discussed later. The sharp posterior border of *Otx2* expression coincides with the anterior expression border of another homeobox gene, *Gbx2*, in the anterior hindbrain, the interface between the domains being stabilized by mutual repression between the two genes (Fig. 13.9A). Experimentally extending the expression domain of *Otx2* into hindbrain territory causes the *Gbx2* border and MHB differentiation to retreat posteriorly, whereas extending the *Gbx2* domain into midbrain territory results in a corresponding anterior shift of the *Otx2* border (Rhinn & Brand, 2001).

Summary

Gene expression patterns have illuminated the time course and mechanisms underlying neural patterning. Functional studies show that homeobox genes direct AP pattern formation, establishing head-to-tail regionalization. In the hindbrain, *Hox* genes encode subregional (rhombomere) identity. At the midbrain–hindbrain boundary, *Otx*

and *Gbx* genes set up and position the MHB organizer, and *En* genes induced at the MHB confer AP polarity on the optic tectum.

Organizers of AP Pattern

Homeobox genes and other classes of transcriptional mediator regulate position-specific development. How is the expression of these genes, or their upstream regulators, directed to specific domains of the AP axis? In many cases, it appears that signals from local organizers are involved, directly or indirectly, in activating specific regional control elements of these genes at appropriate levels of the axis. However, in the case of *Hox* gene regulation, a more distributed source of retinoic acid (RA) signaling has this role of positional signal.

RA exerts its effects via multiple types of RA receptors, which are ligand-dependent transcription factors that bind to specific response elements in the promoters of target genes, including those of the *Hox* genes. In the early embryo, RA is produced by the somites that lie alongside the caudal hindbrain and spinal cord, through the activity of Raldh2. Expressed anterior to the hindbrain is another enzyme, Cyp26, which degrades RA. These enzymes appear to act, respectively, as source and sink for a gradient of RA activity that traverses the AP length of the hindbrain (White, Nie, Lander, & Schilling, 2007). Thus, targeted mutation of *Raldh2* in mice and treatment of chick embryos with an antagonist that specifically blocks all RA receptors both cause anteriorization of the hindbrain.

Immediately anterior to the expression domain of *Hox* genes is a large territory, r1 and midbrain, that forms the cerebellum and optic tectum. A unitary process of specification for these adjacent regions is reflected both at the molecular level and by aspects of their developmental potential: the *Drosophila engrailed* homologues, *En1* and *En2* and the *paired* homologues, *Pax2*, *Pax5*, and *Pax8*, are expressed in this domain. The development of midbrain and r1 is coordinated by signals from the MHB region itself: heterotopic grafts of MHB cells locally induce *En* expression in the host and change the fate of the host neuroepithelium such that it ultimately forms tectal structures (when grafted to the posterior diencephalon) or cerebellar structures (when grafted to the posterior hindbrain). These findings provide compelling evidence that a signal emanating from the MHB is involved in local AP patterning.

There are two prominent components of this organizing activity: Wnt1, expressed in a transverse ring just anterior to the MHB, acts as a mitogen and to maintain the expression of *En* but is unable to mimic the activity of MHB grafts. FGF8, expressed immediately posterior to the *Wnt1* domain, has both inducing and polarizing abilities (Figs. 13.9A and 13.10). When a bead coated with recombinant FGF8 is implanted in the posterior diencephalon of chick embryos, *Fgf8*, *Wnt1*, and *En2* expression is induced in the surrounding cells. The posterior diencephalon then becomes transformed into midbrain, whose AP polarity is reversed with respect to that of the normal host midbrain in accord with the gradient of induced *En* expression (Crossley et al., 1996). Loss of FGF8 function in the zebrafish mutant *Ace* results in loss of the MHB, cerebellum, and part of the tectum. Thus, *Fgf8* expression at the MHB is both necessary and sufficient to establish the polarized pattern of adjoining regions. The asymmetric responses of the mid- and hindbrain anlagen to the FGF8 signal is at least partially due to the prior expression of a competence factor: *Irx2* (a homologue of *Drosophila* prepattern gene *Iroquois*) is expressed early in presumptive r1 and directs the tissue to respond to FGF8 by making cerebellum. Ectopic expression of Irx2 in the midbrain transforms presumptive optic tectum into cerebellum on reception of the FGF8 signal, whereas expressing a constitutive repressor form of Irx2 in r1 transforms presumptive cerebellum into optic tectum. Interestingly, the role of FGF8 in this mechanism is to convert Irx2 protein from a repressor into an activator, which it does by phosphorylation via the MAP kinase pathway (Matsumoto et al., 2004). Diffusion of FGF8 away from the MHB (Scholpp & Brand, 2004) can be seen to act as an on-switch, setting the timing and extent of development of regions whose final identity has already been set up in latent form by prepattern gene expression.

Two further local organizers operate in more rostral regions of the neural tube where a number of early expressed developmental control genes are expressed in a patchwork quilt of small domains, subdividing the diencephalon and telencephalon (the main posterior and anterior subdivisions of the forebrain) into a number of transverse and longitudinal domains (Fig. 13.9). Signaling at the anterior end of the axis involves FGF secretion from a region known as the anterior neural ridge (ANR), which is responsible for inducing and/or maintaining telencephalic (e.g., *Foxg1)* gene expression and for maintaining telencephalic identity (Rubenstein & Beachy, 1998). The ANR is probably derived from the ANB cells that earlier in development secrete antagonists to Wnts (Houart et al., 1998), whose posteriorizing activity must be reduced in order to establish anterior neural identity, as discussed earlier.

A second local organizer of forebrain pattern, the zona limitans intrathalamica (ZLI), lies at the middle of the prospective diencephalon, roughly midway between the ANR and MHB, and is thought to be induced at the interface between Fez and Irx3 expression domains (Fig. 13.9A, B). The ZLI secretes the signal molecule Sonic Hedgehog (Shh), which is responsible for inducing the expression of regional determinant genes both anteriorly (e.g., *Dlx2* in

the prethalmus) and posteriorly (e.g., *Sox14*, *Gbx2* in the thalamus). Like the action of FGF8 at the MHB, the different inductive responses to Shh signaling from the ZLI are due, at least in part, to prepatterned differences in competence of the flanking regions: the *Iroquois* gene *Irx3* is expressed only posterior to the ZLI and directs thalamic differentiation on receipt of the Shh signal (Kiecker & Lumsden, 2005, 2012). The activity of Shh from the ZLI has an apparent dose-dependent activity on the thalamic primordium—cells close to the source, at a presumed high concentration of morphogen, express *Sox14* and develop into GABAergic neurons, whereas those further away express Gbx2 and later form glutamatergic neurons (Jeong et al., 2011; Scholpp et al., 2009; Vue et al., 2009).

Some Key Characteristics of Local Organizers

One of the hallmarks of hindbrain segmentation is the prevention of cell mixing between adjacent rhombomeres. This cell-lineage restriction phenomenon also exists at other boundaries that coincide with organizers (Kiecker & Lumsden, 2005): cell lineage restriction at the MHB has been demonstrated by sophisticated time-lapse imaging in zebrafish embryos and genetic fate mapping in the mouse (Langenberg & Brand, 2005; Zervas, Millet, Ahn, & Joyner, 2004), whereas the ZLI is flanked on both sides by lineage restriction boundaries (Zeltser, Larsen, & Lumsden, 2001). Lineage restriction has not been tested at the ANB, but it seems unlikely that cells intermingle freely across the neural-epidermal border.

Cell lineage restriction minimizes the contact between flanking cell populations, the interface between them, thus tending to form a sharply defined and often straight line across the region. The configuration and positional stability conferred by cell lineage restriction may be seen generally as a requirement for local organizers, ensuring that they can produce stereotypic and predictable molecular gradients in adjacent tissues (Irvine & Rauskolb, 2001).

A further peculiarity of the signal-producing cells in local organizer boundaries is that they proliferate more slowly and differentiate into neurons much later than those of flanking regions (Guthrie, Butcher, & Lumsden, 1991; Lumsden & Keynes, 1989). Transcription factors of the Hes family that mediate Notch signaling are required to inhibit neurogenesis at the MHB of zebrafish and frog embryos (Geling, Plessy, Rastegar, Strähle, & Bally-Cuif, 2004; Ninkovic et al., 2005; Takada, Hattori, Kitayama, Ueno, & Taira, 2005). All neural progenitors express *Hes* genes, but they usually become downregulated when cells undergo neurogenesis. An analysis in mouse embryos has revealed that in boundary regions *Hes* genes remain expressed and that it is this strong persistent expression that sets boundaries apart and allows organizer regions to persist (Baek, Hatakeyama, Sakamoto, Ohtsuke, & Kageyama, 2006).

Summary

Local signaling mechanisms direct the spatial expression of homeobox genes, which appear to be crucial for the delineation and subsequent development of specific anteroposterior subregions of the CNS. Retinoic acid, diffusing from the cervical somites, is in part responsible for establishing the nested expression domains of the *Hox* homeobox genes in the hindbrain. FGF8, produced by cells at the midbrain/hindbrain boundary (MHB), is involved in the elaboration of midbrain and cerebellar structures. Shh, produced by cells at the zona limitans intrathalamica (ZLI), is responsible for elaborating prethalamic and thalamic organization. For both the MHB and the ZLI, tissues lying anterior and posterior to the signaling center respond differently to the secreted morphogen because they have different competences, endowed at least in part by the prior expression of *Irx* prepattern genes. Local organizers are characterized by slow proliferation and late or absent neurogenesis.

AP Pattern of Spinal Nerves

Spinal nerves form a ladder-like array on either side of the spinal cord, an obvious manifestation of segmentation, but here, in contrast to the intrinsic patterning mechanism that operates in the hindbrain, spinal segmentation is imposed extrinsically by the somatic mesoderm that lies alongside. A serially reiterated asymmetry in the sclerotomal component of the somites allows axon to grow, and neural crest cells to migrate, only within the anterior half of each somite, where neural crest cells then condense to form dorsal root ganglia (DRG). The posterior half of each somite expresses glycoproteins that inhibit cell migration and cause growth cone repulsion or collapse. Thus, microsurgically substituting posterior halves for anterior halves results in the local absence of peripheral nerves, whereas the converse experiment results in enlarged DRG and motor nerves. Subdivision of the paraxial mesoderm into AP-polarized somites thus ensures a positional correspondence between the segmented dermomyotome on the one hand and its sensorimotor innervation on the other (Fig. 13.11).

Although it lacks intrinsic segmentation and has a superficial uniformity of organization, the developing spinal cord does manifest distinct AP variations in cellular subtype composition, particularly with respect to motor neurons arranged in discontinuous longitudinal columns. Posterior *Hox* genes have domains of expression in the spinal region that underlie this regional cellular diversity (Dasen & Jessell, 2009; Tsuchida et al., 1994).

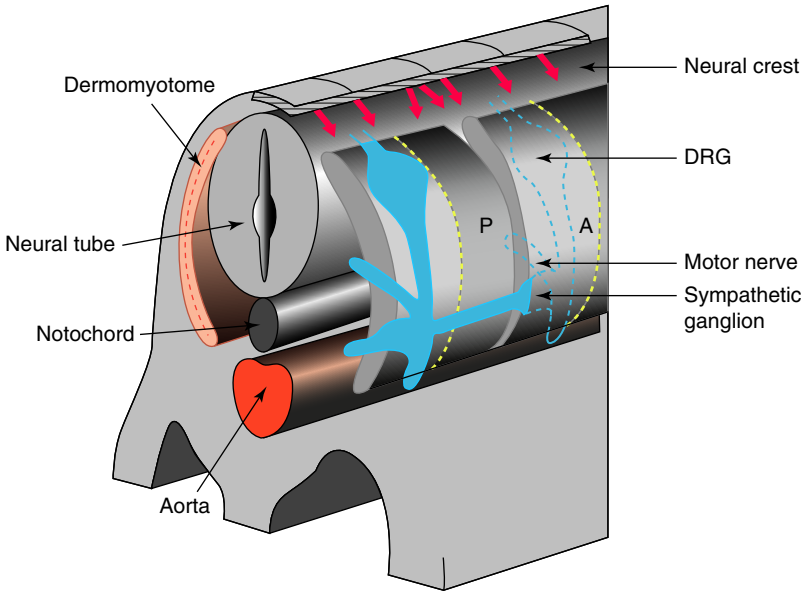


FIGURE 13.11 Segmented pathways for motor axon growth and neural crest migration in the trunk region. The AP positions at which motor axons collect to form a ventral root motor nerve and crest cells aggregate to form a dorsal root ganglion (DRG) are determined by the somites. By the stage at which these constituents of the peripheral nervous system appear, the medial (sclerotomal) region of somites has dispersed and surrounds the notochord and ventral neural tube. The sclerotome is divided into anterior (light gray) and posterior (dark gray) halves, which are distinct from each other according to cell density and molecular markers. All components of the peripheral nervous system are confined to the anterior half sclerotome of each somite, which is permissive for the migration of neural crest cells (arrows), some of which condense to form DRG, and the ingrowth of motor neuron growth cones.

Local Organizers of DV Pattern

The entire neural tube has a characteristic DV zonation where different cell types differentiate stereotypically at different DV positions. The ventral midline of the spinal cord, hindbrain, midbrain, and posterior diencephalon is formed by a palisade of specialized glia, the floor plate. Above the floor plate at spinal levels are five zones containing, in ventral to dorsal sequence, V3 interneurons, motor neurons, and three other types of interneuron, V2, V1, and V0 (Fig. 13.13A). Further subtypes of interneuron differentiate in the dorsal half, whereas the dorsal-most region, represented early on by neural folds that mark the transition between cells with neural and epidermal fates, forms the migratory neural crest cells that give rise to the glia and the majority of neurons in the peripheral nervous system. Later, after the neural crest has departed, the dorsal midline is populated by nonneurogenic roof plate cells.

The patterning of neuronal cell types has been studied intensively in the spinal neural tube, but the same molecules appear to be involved in regionalizing the DV axis of more anterior regions of the neuraxis. Unlike the AP axis, where local organizers signal position and fate to neighboring cells through single morphogen gradients, the DV axis has organizers at both dorsal and ventral poles, and their respective morphogen gradients appear to counteract one another.

Ventral Organizers

Crucially involved in patterning the ventral neural tube is the notochord, a narrow rod that forms from the AME and occupies the midline of the embryo directly beneath the neurectoderm. Grafting experiments in avian embryos have shown that both floor plate and motor neuron differentiation depend on notochord signals. Early removal of the notochord results in a normal-sized spinal cord in which both of these ventral cell types are absent, with dorsal cell types and dorsal-specific markers appearing in their place. Similarly, implanting a supernumerary notochord alongside and in contact with the lateral neural plate results in formation of an additional group of floor plate cells at the point of contact, with clusters of motor neurons on either side (Fig. 13.12A). These experiments show not only the power of the ventral midline signal to influence fate choice but also the multipotent competence of responding neural tube cells at different DV positions. At a slightly later developmental stage, the floor plate itself acquires the same inductive capabilities; it can also induce motor neurons and will induce itself homeogenetically. The floor plate thus becomes an organizing center for a ventral pattern that is built into the neural tube itself.

Although it first appeared that motor neuron and floor plate induction would require different signals, one diffusible and the other contact dependent, a single molecule can account for both processes (for a review, see Jessell, 2000). *Sonic hedgehog (Shh)* is expressed first in the notochord and then in the floor plate and can elicit ectopic floor plate and motor neuron differentiation when misexpressed in the dorsal neural tube. Similarly, an activated form of Smoothened, the signal-transducing component of the Shh receptor, acts in cell autonomous fashion to produce ventral cell types ectopically. Explant studies have shown that the choice of cell fate by progenitor cells is influenced by the concentration of the diffusible Shh protein to which they are exposed: threefold incremental

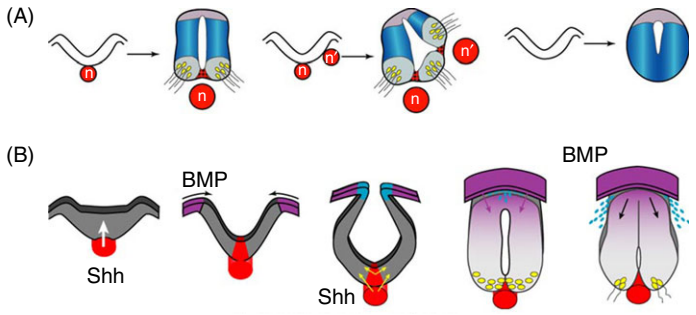


FIGURE 13.12 (A) Influence of ventral midline signals on spinal cord pattern. Cross sections through the developing chick spinal cord at the neural plate stage and resulting spinal cord, showing the effect of adding or removing notochord. (Left) Normal development: the floor plate (red) develops above the notochord (n) and motor neurons (yellow) differentiate in the adjacent ventrolateral region of the neural tube; *Pax6* (blue) is expressed in more dorsal regions. (Center) Grafting a donor notochord (n′) alongside the folding neural plate results in formation of an additional floor plate and a third column of motor neurons. *Pax6* expression retreats from the transformed region. (Right) Removing the notochord from beneath the neural plate results in the permanent absence of both floor plate and motor neurons in the region of the extirpation. *Pax6* expression extends through the ventral region of the cord. (B) Sequence of stages in the formation of DV pattern in the spinal cord and hindbrain. Notochord-derived Shh protein (red) induces differentiation of the floor plate, which also expresses *Shh*. BMPs (dark blue) are expressed in epidermal ectoderm adjoining the neural plate. As the neural plate closes, neural crest cells individuate (light blue) at the junction between neural and epidermal ectoderm. At the early neural tube stage, *Isl1*–expressing motor neuron precursors (yellow) appear close to the floor plate and neural crest cells leave the dorsal tube and midline ectoderm through breaks in the basal lamina. *BMP* expression transfers to the dorsal neural tube Finally, motor neurons differentiate in the ventral cord. White and yellow arrows denote Shh signaling; black and purple arrows denote Bmp signaling.

changes in Shh concentration can result in the generation of five distinct classes of ventral neural tube cell types *in vitro*, including, at the higher end of the range, motor neurons, V3 interneurons, and floor plate cells (Fig. 13.13B). Midventral neural plate cells that are contacted by the notochord are thus likely to be exposed to a high local concentration of Shh, exceeding the threshold for floor plate induction, whereas the lower levels of Shh that diffuse from floor plate are sufficient to induce motor neuron and ventral interneuron differentiation, but insufficient to induce the floor plate. Antibodies that block Shh activity abolish the induction of ventral cell types, showing that Shh is necessary as well as sufficient for establishing the ventral polarity of the neural tube. Importantly, the presumed declining ventral-to-dorsal gradient of Shh protein has been demonstrated in mouse embryos engineered to express a fluorescent-labelled form of Shh (Shh:GFP) from the *Shh* locus (Chamberlain, Jeong, Guo, Allen, & McMahon, 2008).

The ability of midline signals to influence development of the ventral neural tube is not restricted to the spinal cord. In addition to inducing motor neurons and floor plate in the hindbrain and midbrain, midline signals also are involved in the development of AP region-specific neuronal subpopulations. Serotonergic neurons of the hindbrain raphé nucleus form close to the hindbrain floor plate and dopaminergic neurons of the substantia nigra develop within the floor plate (Ono et al., 2007). Both cell types can be induced to form in competent neuroepithelium by the notochord, floor plate, or Shh protein. The issue of how ventral midline signaling can elicit different responses at different AP positions is considered later.

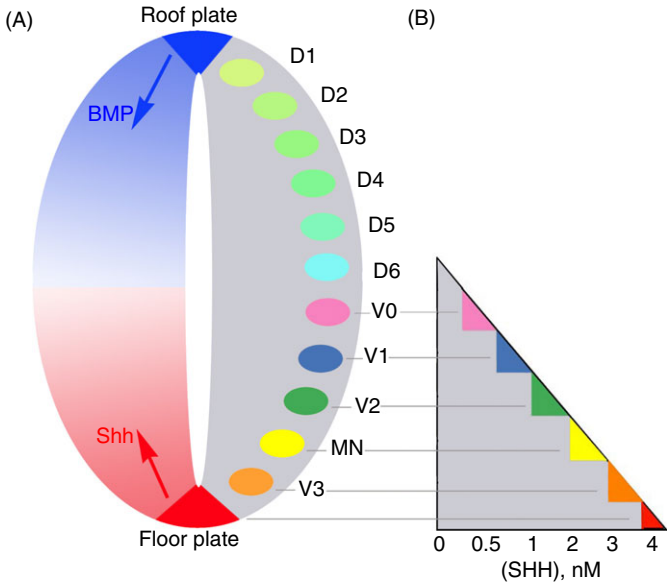


FIGURE 13.13 (A) Formation of different cell types at different DV positions in the spinal cord. Sonic hedgehog (SHH) is responsible for generating interneuron (V0–V3) and motor neuron (MN) diversity in the ventral half of the spinal cord, whereas BMPs are responsible for interneuron (D1–D6) diversity in the dorsal half. SHH protein (red), produced by the floor plate, and BMP protein (blue) by the roof plate act as counteracting graded signals, inducing different cell types at different DV positions *in vivo*. (B) Recombinant SHH protein has been shown to induce a similar range of responses and at different concentrations *in vitro*.

Dorsal Organizers

In genetically or surgically notochordless animals, dorsal markers are expressed in the ventral spinal cord, showing that cell pattern in the dorsal half of the spinal cord

does not require notochord signals. However, dorsal cell types such as neural crest and roof plate, do not develop by default, but are induced by an interaction that initially takes place between the neurectoderm and the flanking epidermal ectoderm when the two epithelia are still contiguous (Fig. 13.12B). This dorsalizing activity appears to be mediated by multiple types of Wnt and BMP proteins, which are expressed in the dorsal ectoderm and can mimic the ability of epidermal ectoderm to induce neural crest cells and roof plate differentiation. As for the ventral neural tube, where initial Shh activation is passed on to the floor plate by homeogenetic induction, so for the dorsal neural tube where BMPs are subsequently expressed in the roof plate and dorsal neural tube itself. BMP signaling is then responsible for inducing the six types of interneuron (dI1-6) that differentiate at progressively more ventral positions in the dorsal half of the cord (Fig. 13.13A; Helms & Johnson, 2003; Liu & Niswander, 2005).

The DV axis is thus patterned by two counteracting gradients, emanating from the ventral and dorsal poles. At either pole, the transfer of signal molecule expression from notochord or epidermal ectoderm to the neuroepithelium attends the physical separation of the neural tube from the initial source, presumably ensuring a more precise control of concentration within the tube.

Definition of DV Pattern by Differential Homeobox Gene Expression

At the time of neuronal differentiation, each of the five zones of neuronal cell types in the ventral neural tube is defined by the combinatorial expression of homeobox genes. These genes are either induced (*Nkx2.2*, *Nkx6.1*) or repressed (*Pax6*, *Pax7*, *Dbx1*, *Dbx2*) at defined concentrations of Shh protein (Fig. 13.14). The differential responses to a graded Shh signal are thought to set up a pattern of fuzzy zones in the ventral half of the cord, which is later refined and sharpened by cross-repressive interactions between the homeobox genes themselves (Briscoe & Ericson, 2001): the expression borders of complementary genes are seen to move following gain and loss of function of the opposing gene, as for *Otx2/Gbx2* at the MHB. Such shifts in the expression domains of these and other homeobox genes result in corresponding and predictable changes in neuronal identity, demonstrating the critical role of these homeobox genes as determinants of cell fate.

As for the AP axis, where positional determinants of identity are conserved with *Drosophila*, there are also striking similarities for the DV axis. For example, the homologue of *Nkx2.2*, *vnd*, is expressed close to the ventral midline of the fly embryo and is required for establishing the identity of its ventral-most neuroblasts (Cornell & von Ohlen, 2000).

Summary

Patterned cell differentiation along the DV axis of the neural tube involves the initial medial–lateral polarization of the neural plate and the later generation of distinct cell types at different DV positions under the spatial control and coordinate actions of an SHH-mediated ventralizing signal from the notochord (later from the floor plate) and BMP-mediated dorsalizing signals from the epidermal ectoderm (later from the roof plate). SHH is both necessary and sufficient for inducing a range of ventral cell types, including floor plate cells, motor neurons, and interneurons, at different concentrations; it therefore has the prime characteristic of a morphogen. Patterning along the entire DV axis involves the intersection of opposing signals emanating from the two poles: BMPs act to limit the ventralizing activity of SHH, whereas SHH acts to limit the dorsalizing activity of BMPs.

Intersection of AP and DV Patterning Mechanisms

The early inductive signals that establish DV cell fate appear to be similar along the entire epichordal neuraxis.

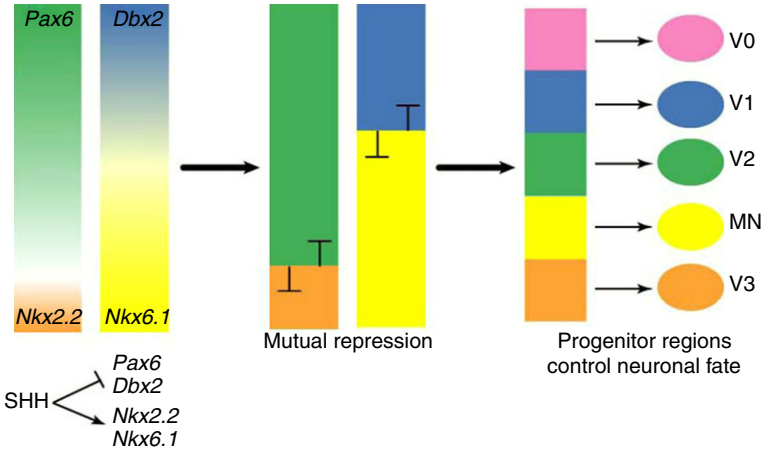


FIGURE 13.14 Model for ventral neural patterning. (Left) Graded SHH signaling from the ventral pole induces new expression of some homeobox genes (e.g., *Nkx2.2, Nkx6.1)* and represses the existing expression of others (e.g., *Pax6, Dbx2*). (Center) Cross-repressive interactions between pairs of transcription factors sharpen mutually exclusive expression domains. (Right) Profiles of homeobox gene expression define progenitor zones and control neuronal fate. *From Briscoe and Ericson (2001).*

However, at a same DV position, there are marked differences in the identity of neurons at different AP positions, such as between oculomotor neurons and dopaminergic neurons of the midbrain and branchiomotor neurons of the hindbrain. That these distinct cell types arise in response to apparently similar levels of Shh signaling suggests that the choice of fate depends on the competence of the responding tissue. Shh must act within the context of previously established AP positional cues and previously specified AP positional values (Simon, Hornbruch, & Lumsden, 1995).

DV Pattern in the Telencephalon

The DV axis of the telencephalon is conspicuously subdivided into a dorsal pallial region, which becomes the neocortex and archicortex (hippocampus) in mammals, and a ventral subpallial region that forms the basal ganglia. While the pallium remains a sheet-like roof over the lateral ventricles, proliferation in the subpallial region thickens the wall considerably, forming two swellings from which the basal ganglia arise: the lateral ganglionic eminence (LGE), which gives rise to the striatum, and the medial ganglionic eminence (MGE), which produces the globus pallidus. The pallial and subpallial moieties of the telencephalic vesicle are segregated by a longitudinal boundary, which marks the domain border of several dorsoventrally expressed developmental control genes and which may also act as a lineage restriction boundary, at least during early stages of development (Fig. 13.15; Campbell, 2003; Schuurmans & Guillemot 2002).

The telencephalon is devoid of peripherally extending motor neurons, has no floor plate, and is not underlain by notochord. The absence of these ventral and midline structures raises the question of how the bilateral organization of the forebrain and the differentiation of its ventral cell types are controlled. It appears, however, that even in this terminal expansion of the central nervous system a common mechanism is used for ventral patterning. SHH expression spreads rostrally in planar fashion from the diencephalic floor region into the ventral telencephalon, where it induces expression of the transcriptional mediator *Nkx2.1*, characteristic of the MGE (Fuccillo, Joyner, & Fishell, 2006). SHH from the MGE is subsequently required for generation of the more lateral region, the LGE, though some LGE markers, such as *Gsh* are still expressed in *Shh* null mutants (Khotz, Baker, Corte, & Fishell, 1998). The different responses of medial versus lateral regions appear to reflect a change in the competence of telencephalic cells with time, with the resulting DV pattern reflecting the delay in reception of the SHH signal rather than a difference in SHH concentration. BMP and WNT signals emanating from the roof plate and medial pallium regulate tissue growth and the expression of dorsal telencephalic markers such as *Lhx2*, *Emx1*, and *Emx2* (Campbell, 2003).

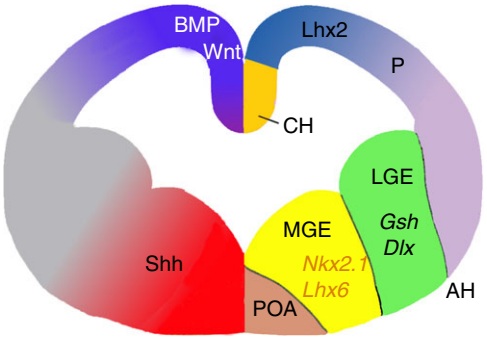


FIGURE 13.15 Cross section of the telencephalon of an early mouse embryo, with signaling on the left side with DV regions and gene expression on the right side. Ventromedial patterning of medial and lateral ganglionic eminences (MGE, LGE) is influenced by SHH signaling from the ventral midline. BMPs and WNTs, produced by dorsomedial regions (roof plate, and medial pallium) and later from the cortical hem (CH), are implicated in pallium (P) specification. AH, anti-hem; POA, preoptic area.

During early forebrain development, the telencephalic roof plate sinks into the prosencepahlic lumen, creating bilateral cortical hemispheres and giving rise both to the choroid plexus medially and to the cortical hem bilaterally. Immediately adjacent to the hem (Fig. 13.15), which expresses BMPs and Wnts, the pallium forms hippocampus and the cerebral cortex forms lateral to that. Genetic ablation of the telencephalic roof plate in mice results in a severe undergrowth of the cortical primordium, suggesting a nonautonomous effect of the hem on cortical specification and proliferation (Monuki, Porter, & Walsh, 2001).

Wnt signaling specifies dorsal identity at the earliest stages of telencephalic development (Gunhaga et al., 2003) and at later stages, Wnt secretion from the cortical hem is required specifically for hippocampus differentiation (Galceran, Miyashita-Lin, Devaney, Rubenstein, & Grosschedl, 2000). The boundary between the pallium and the subpallium (PSB) expresses Wnt inhibitors (secreted Frizzled-related proteins, SFRPs), raising the possibility that a gradient of Wnt activity is established across the pallium between the hem and the PSB (which has also been called the "anti-hem") (Assimacopoulos, Grove, & Ragsdale, 2003; Frowein, Campbell, Götz, 2002). The homeobox gene *Lhx2*, expressed in the pallium between the hem and anti-hem, in the large region destined to become neocortex, has been shown to act as a cell-autonomous determinant of cortical fate: when *Lhx2* is mutated in mice (Fig. 13.16), the *Wnt2b*-expressing hem and the *Dbx1/SFRP*-expressing anti-hem expand in place of neocortex (Mangale et al., 2008).

Patterning the Cortex

The adult mammalian neocortex comprises a very large number of anatomically and functionally discrete areas—more than 50 in humans. One difficulty in addressing the

question of how the elaborate areal pattern of the neocortex develops is that it is virtually uniform in structure and cytoarchitecture from one end to the other, at least in lower mammals such as the mouse and rat. Furthermore, few known developmental control genes were known to be expressed early enough to be candidate determinants for each, or indeed any of these prospective areas. Thus, it was long thought that the areal pattern must be imposed on the neocortex by extrinsic elements, principally via the thalamic afferents that penetrate the cortex in an area-specific manner. However, advances in gene discovery now reveal that the cortical neuroepithelium acquires regional pattern prior to afferent innervation and in a manner not unlike other regions of the neural tube, with transcriptional mediators being expressed in a graded fashion in response to local concentrations of morphogens diffusing from focal sources. The latter include the ANR, which secretes FGFs (see above, and Toyoda et al., 2010), and the Wnt-secreting cortical hem. It is likely that areal specification involves threshold levels of transcriptional control genes such as *Emx2* and *Pax6*, which are expressed in opposing gradients along the AP extent of the cortex (Fig. 13.17), and others such as *Lef1* (a mediator of Wnt signaling) that display graded expression from medial to lateral. In support of this, loss of *Emx2* function shrinks visual cortex (posterior) and expands sensory (mid) and motor (anterior) areas, whereas loss of *Pax6* function results in the opposite alteration. Furthermore, an ectopic source of FGF8 at the posterior end of the cortex results in a partial mirrored duplication of anterior cortical areas (Fukushi-Shimogori & Grove, 2001). Thus, in contradiction to longstanding ideas, cortical pattern is probably set up early in the parent neuroepithelium and the role of thalamic innervation is in secondary reinforcing, refining, or modifying of this initial pattern (O'Leary & Nakagawa, 2002; Ragsdale & Grove, 2001).

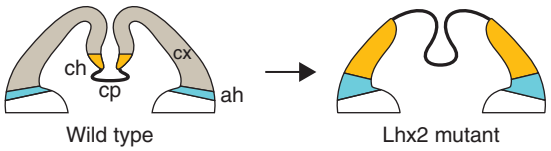


FIGURE 13.16 Schematic showing the effect of *Lhx2* mutation on pallial fate, in cross section through the dorsal telencephalon. When Lhx2 function is ablated, the neocortex (cx) fails to form and its pallial progenitor region instead forms expanded choroid plexus (cp), cortical hem (ch), and anti-hem (ah). *Data from Mangale et al. (2008).*

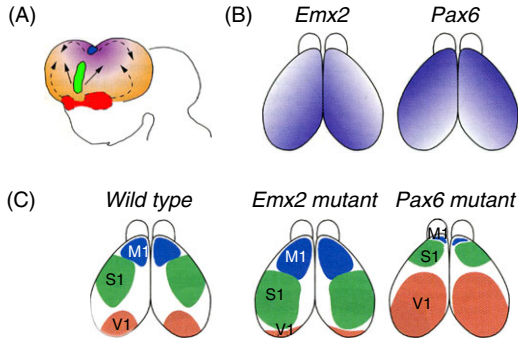


FIGURE 13.17 (A) Regulation of neocortical pattern by local signals: FGF8 (green) from the anterior neural ridge (ANR), Wnts and Bmps (blue) from the cortical hem; and SHH (red) from the medial ganglionic eminence. (B) These and other signals are thought to induce expression of the transcriptional factors *Emx2* and *Pax6* in opposed gradients that control regional identity. (C) Loss of function of either of these genes results in a specific shift of area identity from wild type (left), implying that visual cortex (V1) is specified by *Emx2* [high], *Pax6* [low] and that of motor cortex is specified by *Emx2* [low], *Pax6* [high]. *From O'Leary and Nakagawa (2002).*

CONCLUSIONS

Neural-inducing factors and modifiers produced during gastrulation have a basic role in establishing an initial crude AP regional identity in the neural plate that emerges from the dorsal surface of the embryo toward the end of this period. Abrogation of Wnt signaling, by depletion of ligand or blockade of signal transduction pathways, is important in conferring anterior neural identity. Wnts, Fgfs, and RA cooperate to posteriorize the neuraxis. The resulting coarse-grained AP pattern is subsequently refined by the action of local organizers. Acting on both the AP and DV axes, local organizers exert regional growth control and either induce or repress transcriptional mediators of regional fate. These mediators, in turn, may also interact among themselves, sharpening domains of action by cross-repression. The combined processes of segmentation and compartition provide an alternative or perhaps additional mechanism for sharpening gene expression borders and cellular domains. These events result in the precise delineation of regions and subregions. Although we tend to think of neuraxial patterning as a series of discrete steps leading to greater refinement, it is a continuous process that extends over a protracted period of development and involves the repetitive use of signaling systems (e.g., RA, FGFs, SHH, BMPs, Wnts). We have also presented this subject as if pattern is acquired separately on AP and DV axes, whereas it is clear that cell specification on the DV axis must intersect with and act on already existing AP positional values. It is also clear that patterning mechanisms must be integrated with those that control neurogenesis, a topic that is considered in Chapter 15.

References

Assimacopoulos, S., Grove, E. A., & Ragsdale, C. W. (2003). Identification of a Pax6-dependent epidermal growth factor family signaling source at the lateral edge of the embryonic cerebral cortex. *Journal of Neuroscience, 23,* 6399–6403.

Bachiller, D., Klingensmith, J., Kemp, C., Belo, J. A., Anderson, R. M., May, S. R., … De Robertis, E. M. (2000). The organizer factors Chordin and Noggin are required for mouse forebrain development. *Nature, 403,* 658–661.

Baek, J. H., Hatakeyama, J., Sakamoto, S., Ohtsuke, T., & Kageyama, R. (2006). Persistent and high levels of *Hes1* expression regulate boundary formation in the developing central nervous system. *Development, 133,* 2467–2476.

Baker, J. C., Beddington, R. S., & Harland, R. M. (1999). Wnt signaling in *Xenopus* embryos inhibits *bmp4* expression and activates neural development. *Genes and Development, 13,* 3149–3159.

Barth, K. A., Kishimoto, Y., Rohr, K. B., Seydler, C., Schulte-Merker, S., & Wilson, S. W. (1999). Bmp activity establishes a gradient of positional information throughout the entire neural plate. *Development, 126,* 4977–4987.

Bell, E., Wingate, R., & Lumsden, A. (1999). Homeotic transformation of rhombomere identity after localized *Hoxb1* misexpression. *Science, 284,* 2168–2171.

Bier, E. (1997). Anti-neural-inhibition: A conserved mechanism for neural induction. *Cell, 89,* 681–684.

Briscoe, J., & Ericson, J. (2001). Specification of neuronal fates in the ventral neural tube. *Current Opinion in Neurobiology, 11,* 43–49.

Campbell, K. (2003). Dorsal-ventral patterning in the mammalian telencephalon. *Current Opinion in Neurobiology, 13,* 50–56.

Chamberlain, C. E., Jeong, J., Guo, C., Allen, B. L., & McMahon, A. P. (2008). Notochord-derived Shh concentrates in close association with the apically positioned basal body in neural target cells and forms a dynamic gradient during neural patterning. *Development, 135,* 1097–1106.

Cornell, R. A., & von Ohlen, T. (2000). Vnd/nkx, ind/gsh, and msh/msx: Conserved regulators of dorsoventral neural patterning? *Current Opinion in Neurobiology, 10,* 63–71.

Crossley, P. H., Martinez, S., & Martin, G. R. (1996). Midbrain development induced by Fgf8 in the chick embryo. *Nature, 380,* 66–68.

Dasen, J. H., & Jessell, T. M. (2009). Hox networks and the origins of motor neuron diversity. *Current Topics in Developmental Biology, 88,* 169–200.

DeRobertis, E. M. (2006). Spemann's organizer and self-regulation in amphibian embryos. *Nature Reviews Molecular Cell Biology, 7,* 265–267.

Diez del Corral, R., & Storey, K. G. (2004). Opposing FGF and retinoid pathways: A signalling switch that controls differentiation and patterning onset in the extending vertebrate body axis. *Bioessay, 26,* 857–869.

Fraser, S. E., Keynes, R., & Lumsden, A. (1990). Segmentation in the chick embryo hindbrain is defined by cell lineage restrictions. *Nature, 344,* 431–435.

Frowein, J., Campbell, K., & Götz, M. (2002). Expression of *Ngn1, Ngn2, Cash1, Gsh2* and *Sfrp2* in the developing chick telencephalon. *Mechanisms of Development, 110,* 249–252.

Fuccillo, M., Joyner, A. L., & Fishell, G. (2006). Morphogen to mitogen: the multiple roles of hedgehog signalling in vertebrate neural development. *Nature Reviews Neuroscience, 7,* 772–783.

Fuentealba, L. C., Eivers, E., Ikeda, A., Hurtado, C., Kuroda, H., Pera, E. M., & De Robertis, E. M. (2007). Integrating patterning signals: Wnt/GSK3 regulates the duration of the BMP/Smad1 signal. *Cell, 131,* 980–993.

Fukushi-Shimogori, T., & Grove, E. (2001). Neorcortex patterning by the secreted signaling molecule FGF8. *Science, 294,* 1071–1074.

Galceran, J., Miyashita-Lin, E. M., Devaney, E., Rubenstein, J. L., & Grosschedl, R. (2000). Hippocampus development and generation of dentate gyrus granule cells is regulated by LEF1. *Development, 127,* 469–482.

Geling, A., Plessy, C., Rastegar, S., Strähle, U., & Bally-Cuif, L. (2004). Her5 acts as a prepattern factor that blocks *neurogenin1* and *coe2* expression upstream of Notch to inhibit neurogenesis at the midbrain–hindbrain boundary. *Development, 131,* 1993–2006.

Grunz, H., & Tacke, L. (1989). Neural differentiation of *Xenopus laevis* ectoderm takes place after disaggregation and delayed reaggregation without inducer. *Cell Differentiation and Development, 28,* 211–217.

Guthrie, S., Butcher, M., & Lumsden, A. (1991). Patterns of cell division and interkinetic nuclear migration in the chick embryo hindbrain. *Journal of Neurobiology, 22,* 742–754.

Helms, A. W., & Johnson, J. (2003). Specification of dorsal spinal cord interneurons. *Current Opinion in Neurobiology, 13,* 42–49.

Hirth, F., Hartmann, B., & Reichert, H. (1998). Homeotic gene action in embryonic brain development of *Drosophila*. *Development, 125,* 1579–1589.

Houart, C., Westerfield, M., & Wilson, S. W. (1998). A small population of anterior cells patterns the forebrain during zebrafish gastrulation. *Nature, 391,* 788–792.

Irvine, K. D., & Rauskolb, C. (2001). Boundaries in development: Formation and function. *Annual Review of Cell and Developmental Biology, 17,* 189–214.

Jeong, Y., Dolson, D. K., Waclaw, R. R., Matise, M. P., Sussel, L., Campbell, K., … Epstein, D. J. (2011). Spatial and temporal requirements for sonic hedgehog in the regulation of thalamic interneuron identity. *Development, 138,* 531–541.

Jessell, T. M. (2000). Neuronal specification in the spinal cord: inductive signals and transcriptional codes. *Nature Reviews Genetics, 1,* 20–29.

Jimenez-Guri, E., Udina, F., Colas, J. F., Sharpe, J., Padrón-Barthe, L., Torres, M., & Pujades, C. (2010). Clonal analysis in mice underlines the importance of rhombomeric boundaries in cell movement restriction during hindbrain segmentation. *PLoS One, 5,* e10112.

Kamiya, D., Banno, S., Sasai, N., Watanabe, K., Kawada, M., Yakura, R., & Sasai, Y. (2011). Intrinsic transition of embryonic stem-cell differentiation into neural progenitors. *Nature, 470,* 503–509.

Keller, R. (2002). Shaping the vertebrate body plan by polarized embryonic cell movements. *Science, 298,* 1950–1954.

Khokha, M. K., Yeh, J., Grammer, T. C., & Harland, R. M. (2005). Depletion of three BMP antagonists from Spemann's organizer leads to a catastrophic loss of dorsal structures. *Developmental Cell, 8,* 401–411.

Khotz, J., Baker, D. P., Corte, G., & Fishell, G. (1998). Regionalization within the mammalian telencephalon is mediated by changes in responsiveness to Sonic hedgehog. *Development, 125,* 5079–5089.

Kiecker, C., & Lumsden, A. (2005). Compartments and their boundaries in vertebrate brain development. *Nature Reviews Neuroscience, 6,* 553–564.

Kiecker, C., & Lumsden, A. (2012). The role of organizers in patterning the nervous system. *Annual Review of Neuroscience* (in press).

Kiecker, C., & Niehrs, C. (2001). A morphogen gradient of Wnt/β-catenin signalling regulates anteroposterior neural patterning in *Xenopus*. *Development, 128,* 4189–4201.

Langenberg, T., & Brand, M. (2005). Lineage restriction maintains a stable organizer cell population at the zebrafish midbrain–hindbrain boundary. *Development, 132,* 3209–3216.

Liu, A., & Niswander, L. A. (2005). Bone morphogenetic protein signalling and vertebrate nervous system development. *Nature Reviews Neuroscience, 6,* 945–954.

Lumsden, A., & Keynes, R. (1989). Segmental patterns of neuronal development in the chick hindbrain. *Nature, 337,* 424–428.

Maden, M. (2007). Retinoic acid in the development, regeneration and maintenance of the nervous system. *Nature Reviews Neuroscience, 8,* 755–765.

Mangale, V. S., Hirokawa, K. E., Satyakim, P. R., Gokulchandran, N., Chikbire, S., Subramanian, L., … Monuki, E. S. (2008). Lhx2 selector activity specifies cortical identity and suppresses hippocampal organizer fate. *Science, 319,* 304–309.

Mason, I. (2007). Initiation to end point: The multiple roles of fibroblast growth factors in neural development. *Nature Reviews Neuroscience, 8,* 583–596.

Mathis, L., Kulesa, P. M., & Fraser, S. E. (2001). FGF receptor signalling is required to maintain neural progenitors during Hensen's node progression. *Nature Cell Biology, 3,* 559–566.

Matsumoto, K., Nishihara, S., Kamimura, M., Shiraishi, T., Otoguro, T., & Uehara, M., … Ogura, T. (2004). The prepattern transcription factor Irx2, a target of the FGF8/MAPK cascade, is involved in cerebellum formation. *Nature Neuroscience, 7,* 605–612.

Monuki, E. S., Porter, F. D., & Walsh, C. A. (2001). Patterning of the dorsal telencephalon and cerebral cortex by a roof plate-Lhx2 pathway. *Neuron, 32,* 591–604.

Muñoz-Sanjuán, I., & Brivanlou, A. H. (2002). Neural induction, the default model and embryonic stem cells. *Nature Reviews Neuroscience, 3,* 271–280.

Niehrs, C. (2004). Regional specific induction by Spemann-Mangold organizer. *Nature Reviews Genetics, 5,* 425–434.

Ninkovic, J., Tallafuss, A., Leucht, C., Topczewski, J., Tannhäuser, B., Solnica-Krezel, L., & Bally-Cuif, L. (2005). Inhibition of neurogenesis at the zebrafish midbrain–hindbrain boundary by the combined and dose-dependent activity of a new *hairy/E(spl)* gene pair. *Development, 132,* 75–88.

O'Leary, D. D. M., & Nakagawa, Y. (2002). Patterning centers, regulatory genes and extrinsic mechanisms controlling arealization of the neocortex. *Current Opinion in Neurobiology, 12,* 14–25.

Ono, Y., Nakatani, T., Sakamoto, Y., Mizuhara, E., Minaki, Y., Kumai, M., … Imai, T. (2007). Differences in neurogenic potential in floor plate cells along an anteroposterior location: Midbrain dopaminergic neurons originate from mesencephalic floor plate cells. *Development, 134,* 3213–3225.

Ragsdale, C. W., & Grove, E. A. (2001). Patterning the mammalian cerebral cortex. *Current Opinion in Neurobiology, 11,* 50–58.

Reversade, B., & De Robertis, E. M. (2005). Regulation of ADMP and BMP2/4/7 at opposite embryonic poles generates a self-regulating morphogenetic field. *Cell, 123,* 1147–1160.

Rhinn, M., & Brand, M. (2001). The midbrain–hindbrain boundary organizer. *Current Opinion in Neurobiology, 11,* 34–42.

Rubenstein, J. L. R., & Beachy, P. A. (1998). Patterning the embryonic forebrain. *Current Opinion in Neurobiology, 8,* 18–26.

Scholpp, S., & Brand, M. (2004). Endocytosis controls spreading and effective signalling range of Fgf8 protein. *Current Biology, 14,* 1834–1841.

Scholpp, S., Delogu, A., Gilthorpe, J., Peukert, D., Schindler, S., & Lumsden, A. (2009). *Her6* regulates the neurogenetic gradient and neuronal identity in the thalamus. *Proceedings of the National Academy of Sciences of the United States of America, 106,* 19895–19900.

Schulte-Merker, S., Lee, K. J., McMahon, A. P., & Hammerschmidt, M. (1997). The zebrafish organizer requires *chordino*. *Nature, 387,* 862–863.

Schuurmans, C., & Guillemot, F. (2002). Molecular mechanisms underlying cell fate specification in the developing telencephalon. *Current Opinion in Neurobiology, 12,* 26–34.

Simeone, A., Puelles, E., & Acampora, D. (2002). The Otx family. *Current Opinion in Genetics and Development, 12,* 409–415.

Simon, H., Hornbruch, A., & Lumsden, A. (1995). Independent assignment of anteroposterior and dorsoventral positional values in the developing chick hindbrain. *Current Biology, 5,* 205–214.

Spemann, H., & Mangold, H. (1924). Über induction von Embryonanlagen durch Implantation Artfremder Organisatoren. *Roux' Archiv für Entwicklungsmechanik, 100,* 599–638.

Stern, C. D. (2006). Neural induction: 10 years after the default model. *Current Opinion in Cell Biology, 18,* 692–697.

Studer, M., Lumsden, A., Ariza-McNaughton, L., Bradley, A., & Krumlauf, R. (1996). Altered segmental identity and abnormal migration of motor neurons in mice lacking Hoxb1. *Nature, 384,* 630–634.

Takada, H., Hattori, D., Kitayama, A., Ueno, N., & Taira, M. (2005). Identification of target genes for the *Xenopus* Hes-related protein XHR1, a prepattern factor specifying the midbrain–hindbrain boundary. *Developmental Biology, 283,* . 253–67.

Toyoda, R., Assimacopoulos, S., Wilcoxon, J., Taylor, A., Feldman, P., Suzuki-Hirano, A., Shimogori, T., & Grove, E. A. (2010). FGF8 acts as a classic diffusible morphogen to pattern the neocortex. *Development, 137,* 3439–3448.

Tsuchida, T., Ensini, M., Morton, S. B., Baldassare, M., Edlund, T., Jessell, T. M., & Pfaff, S. L. (1994). Topographic organization of embryonic motor neurons defined by expression of LIM homeobox genes. *Cell, 79,* 957–970.

Vue, T. Y., Bluske, K., Alishahi, A., Yang, L. L., Koyano-Nakagawa, N., Novitch, B., & Nakagawa, Y. (2009). Sonic hedgehog signaling controls thalamic progenitor identity and nuclei specification in mice. *Journal of Neuroscience, 29,* 4484–97.

Wang, Y., & Nathans, J. (2007). Tissue/planar cell polarity in vertebrates: New insights and new questions. *Development, 134,* 647–658.

Wessely, O., Agius, E., Oelgeschläger, M., Pera, E. M., & De Robertis, E. M. (2001). Neural induction in the absence of mesoderm: *β*-catenin-dependent expression of secreted BMP antagonists at the blastula stage in *Xenopus*. *Developmental Biology, 234,* 161–173.

White, R. J., Nie, Q., Lander, A. D., & Schilling, T. F. (2007). Complex regulation of cyp26a1 creates a robust retinoic acid gradient in the zebrafish embryo. *PLoS Biology, 5,* e304.

Wizenmann, A., & Lumsden, A. (1997). Segregation of rhombomeres by differential cell affinity. *Molecular and Cellular Neuroscience, 9,* 448–459.

Wurst, W., Auerbach, A. B., & Joyner, A. L. (1994). Multiple developmental defects in *Engrailed-1* mutant mice: An early midhindbrain deletion and patterning defects in forelimbs and sternum. *Development, 120,* 2065–2075.

Xu, Q., Mellitzer, G., Robinson, V., & Wilkinson, D. G. (1999). In vivo cell sorting in complementary segmental domains mediated by Eph receptors and ephrins. *Nature, 400,* 267–271.

Zeltser, L. M., Larsen, C. W., & Lumsden, A. (2001). A new developmental compartment in the forebrain regulated by *Lunatic fringe*. *Nature Neuroscience, 4,* 683–684.

Zervas, M., Millet, S., Ahn, S., & Joyner, A. L. (2004). Cell behaviors and genetic lineages of the mesencephalon and rhombomere 1. *Neuron, 43,* 345–357.

Zoltewicz, J. S., & Gerhart, J. C. (1997). The Spemann organizer of *Xenopus* is patterend along its anteroposterior axis at the earliest gastrula stage. *Developmental Biology, 192,* 482–491.

Andrew Lumsden and Clemens Kiecker

CHAPTER

14

Cellular Determination

INTRODUCTION

There are a huge variety of cell types in the nervous system with completely different shapes, patterns of connectivity, neurochemistry, and physiological characteristics. The distribution of different types of neuron follows a highly invariant pattern in a given species. Neurons located at a given position within the nervous system are always the same type. In vertebrates, this is true on the level of populations of neurons (e.g., cells of layer III of the cortical area 17 are glutaminergic pyramidal cells projecting to cortical areas 18 and 19); in many invertebrates, it is true even on the level of individual, uniquely identifiable cells. Experimental studies carried out over the last few decades have provided substantial progress in understanding how the right types of neuron develop in the right place in the brain.

Neurons are specified gradually by a series of fate choices that begin with their distant progenitors who make the first decision between neural and ectodermal fates. Neural progenitors, those dividing cells that are determined to give rise to neurons and glia, arise from a specialized region of the ectoderm called the neurectoderm, known in vertebrates as the neural plate (Fig. 14.1). A highly conserved molecular network, discussed in Chapter 13, specifies the size and position of this region. The neural plate rolls or folds into the neural tube and detaches from the overlying epidermis. The neural tube consists of a continuous pseudostratified neuroepithelium that lines, on its apical surface, an inner lumen, called the ventricle (Fig. 14.1C, D). Neuroepithelial cells of the neural tube begin dividing symmetrically at the apical surface or ventricular zone, producing two more cells that divide (Fig. 14.1C′). After a period of time, during which this doubling process leads to an exponential expansion of cell numbers, these neural progenitors may begin to divide asymmetrically or spin off postmitotic daughters that differentiate and other daughters that divide again. Finally, at the end of neurogenesis, cell divisions lead only to differentiating daughters. The number of cells in any part of the nervous system is thus largely governed by the proliferative potential and division patterns of the neural progenitors.

Once a cell has exited the cell cycle, a newly born precursor must make a number of other decisions: whether to become a neuron or a glial cell, whether to be excitatory or inhibitory, what shape to take, what partners to connect with, and so forth. It is important to understand that much of what we know about this topic comes from work in invertebrate species such as *C. elegans* and especially the fruit fly, *Drosophila melanogaster*. Progenitors of the central nervous system of *Drosophila*, called neuroblasts (the equivalent of neuroepithelial cells in vertebrates), separate from the neurectoderm and move inside the embryo as individual cells (Fig. 14.1G). Mitotic divisions of *Drosophila* neuroblasts are asymmetric throughout development. Thus, each neuroblast divides unequally into one large and one small daughter cell. The large cell remains as a neuroblast and continues to divide asymmetrically for a number of rounds (Fig. 14.1G′). The small cell, called the ganglion mother cell (GMC), typically divides one more time and both of its daughter cells differentiate into mature neurons.

In this chapter, we begin to explore the molecular and cellular aspects of determination, such as the genes that specify neural versus nonneural fate, and how position and time of birth shape a neuron's fate as well as the mode of division of its parental cell. Running through all this is a network of transcription factors that influence almost every aspect of a neuron's specific identity. We will then discuss how these factors might work together and go through several well studied examples of neuronal determination in which we see them at work. These include neurons versus glia, the sensory cells of the peripheral nervous system, the retina, the spinal cord, and finally the cerebral cortex.

© 2013 Elsevier Inc. All rights reserved.

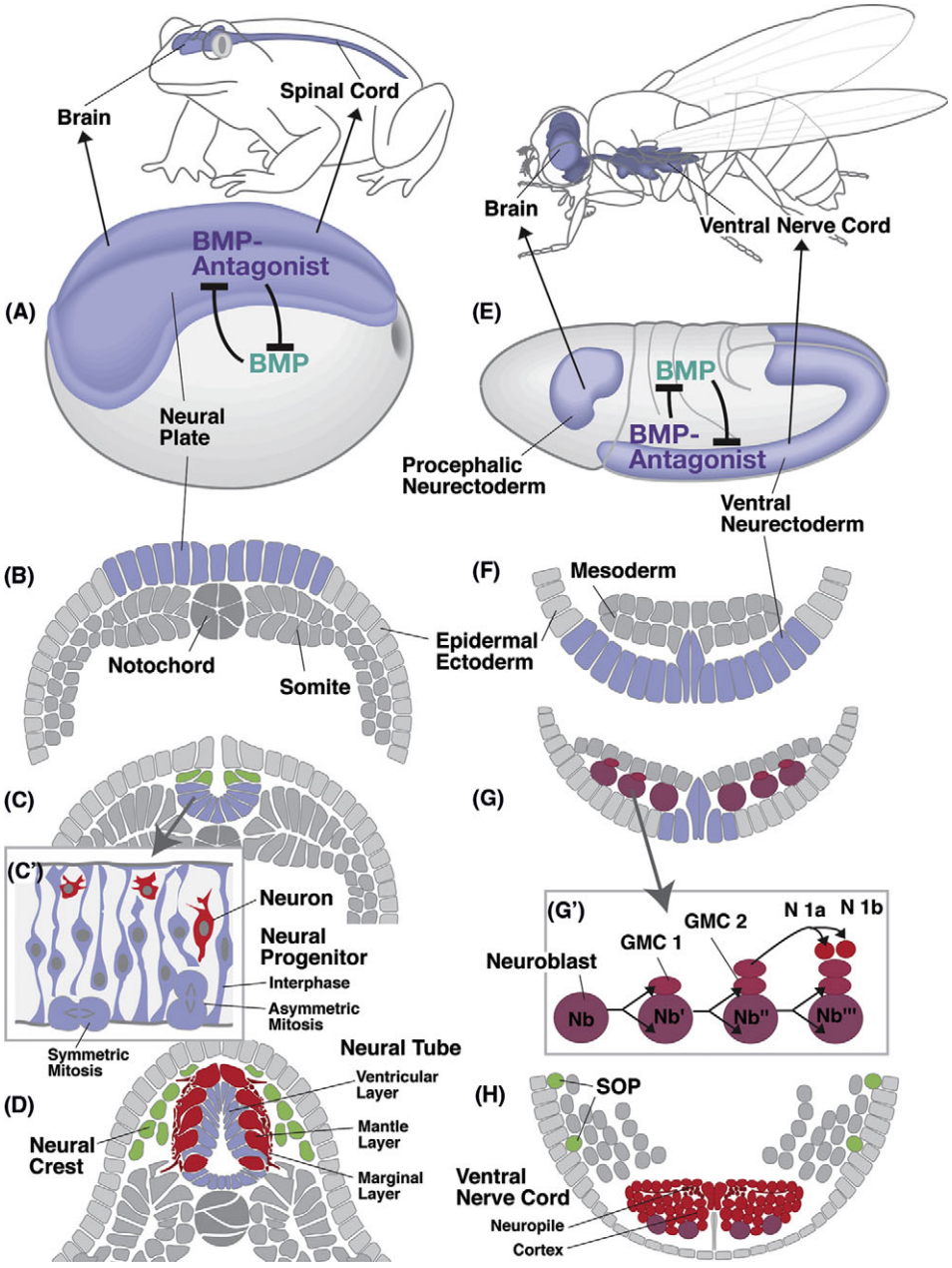


FIGURE 14.1 Synopsis of early neural development in vertebrates (example: the amphibian *Xenopus laevis*; A–D) and insects (example: the fruitfly *Drosophila melanogaster*; E–H). (A and E) Dorsolateral views of embryos at the onset of neurulation. The neurectoderm is shaded light blue. The expression of the signal Bone Morphogenetic Protein (BMP) and its antagonists is indicated. (B and F) Schematic cross sections of embryos depicted in (A and E). (C and G) Cross sections at a later stage when neurulation is well under way. In vertebrates, the neural plate folds in to become the neural tube. Cells at the junction between the neural tube and the epidermal ectoderm (green) form the neural crest, which gives rise to the peripheral and autonomic nervous systems. The neural tube is a proliferating neuroepithelium (C'); symmetric mitoses increase the number of epithelial neural progenitor cells, asymmetric mitoses result in daughter cells that delaminate from the epithelium and become neurons (red). In insects, individual neural progenitors (neuroblasts; purple) delaminate from the neurectoderm. They divide in a stem cell mode (G'), producing stacks of daughter cells called ganglion mother cells (GMC). Each ganglion mother cell divides into two neurons (e.g., N1a and N1b). (D and H) Cross sections of late embryos in which some neurons (red) have differentiated. In vertebrates, these neurons form the mantle layer that surrounds the neuroepithelium, now called ventricular layer. Neurites gather at the outside of the neural tube (marginal layer). In insects neuronal cell bodies form an outer layer (cortex); neurites gather in the center, forming the neuropile. Progenitors of the peripheral nervous system [sensory organ progenitors (SOPs); green] segregate from different locations in the epidermis.

NEUROGENESIS

Neurogenesis proceeds quite differently in different parts of the neural primordium. For example, neuroepithelial cells in the anterior part of the vertebrate neural tube (which includes the forebrain) undergo many more rounds of division than those of the posterior part (spinal cord). In addition, the maturing neurons have distinct phenotypes and establish different types of connections depending on where they are located. When pieces of neural tube are transplanted to new positions, features such as proliferative potential and cellular phenotype are often carried to the new position, implying that, in the neural primordium, intrinsic determinants profoundly influence the fate of neural progenitors. These determinants subdivide the neural primordium into a "geographic map" with many different domains, each one characterized by a unique identity. The mechanisms underlying the spatial regionalization, discussed in more detail in Chapter 13, are intricately involved in neural fate specification.

Along its anterior-posterior axis the neural tube is subdivided into transverse domains that give rise to morphologically defined partitions of the CNS, including the spinal cord, hindbrain, midbrain, and forebrain (Fig. 14.2B). Each of these major sections of the neural tube becomes further

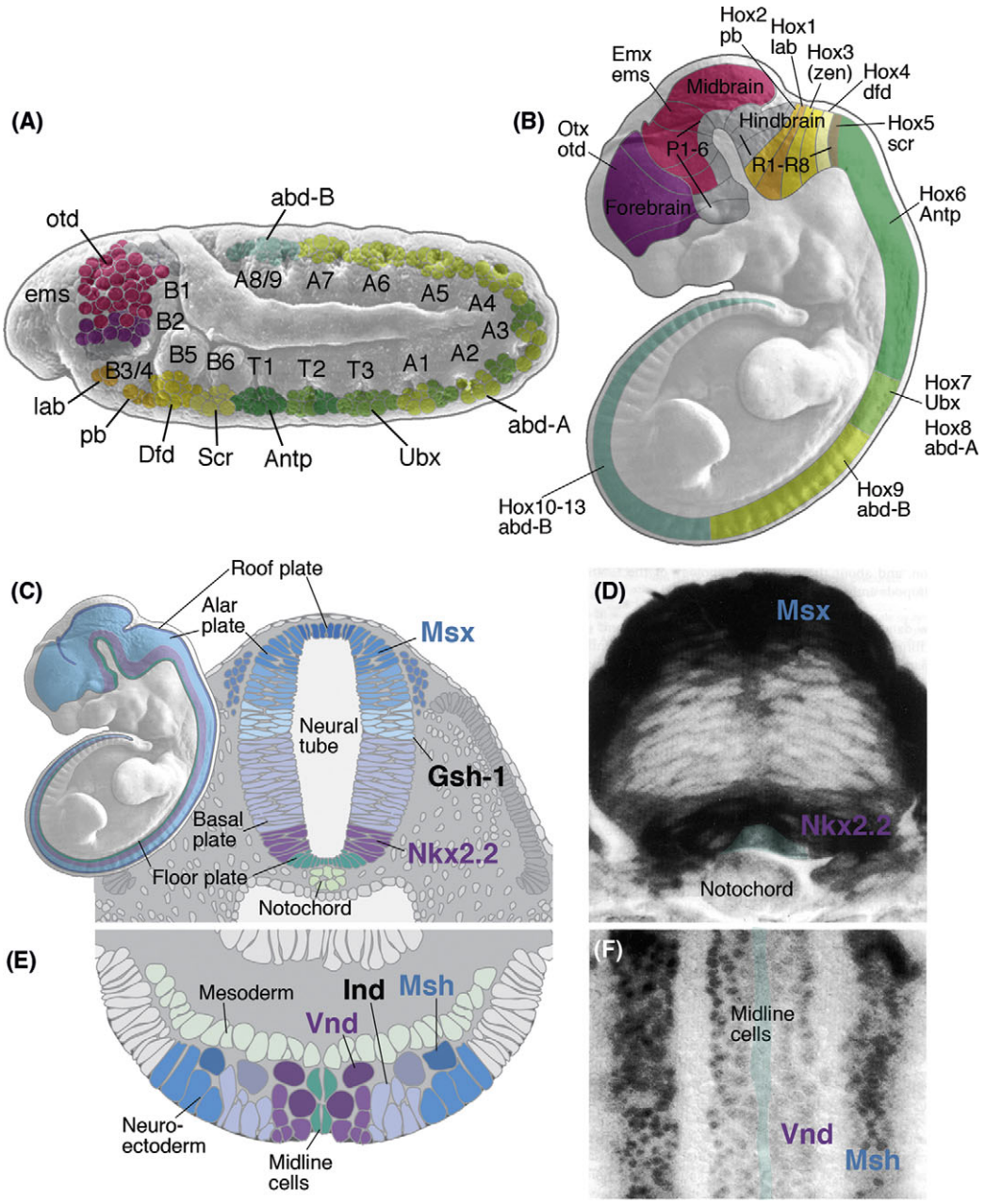


FIGURE 14.2 Specification of regional identity of the neural primordium. (A and B) Homeobox genes control regional identity along the antero-posterior axis. (A) *Drosophila*. The neurectoderm and neuroblast layer is subdivided into a series of neuromeres, each of which gives rise to a segmental ganglion. Transcription factors of the Hox family are expressed in blocks of neuromeres and provide neuroblasts with a segment-specific identity. In the segments of the abdomen (A1–A9), thorax (T1–T3), and posterior head (B3–B6) genes of the Hox complex are expressed (*lab, pb, Dfd, Scr, Antp, Ubx, AbdA, AbdB*); head gap genes (*otd, ems*) are found in the anterior head (B1–2). B: Mouse. Homologous Hox genes and Head gap genes are expressed in sets of vertebrate neuromeres in the same antero-posterior sequence (P1–P6: prosomeres of forebrain; R1–7: rhombomeres of hindbrain). Otx is expressed in the dorsal telencephalon (P1–3) that gives rise to cerebral cortex. Emx defines alar plate of prosomeres 4–6, that will form the dorsal thalamus and ventral thalamus, as well as the dorsal midbrain. Hox genes are expressed in hindbrain and spinal cord. Note that for the Hox genes, the level indicated represents the anterior boundary of a wider domain that extends posteriorly for an unspecified distance. (C–F) Regional identity along the dorso-ventral axis. (C and D) Vertebrate. The neural tube (shown in C for mouse) is subdivided into four domains, the floor plate, basal plate, alar plate, and roof plate. Floor plate and roof plate do not give rise to neurons. The basal plate produces motoneurons and interneurons. Interneurons formed by the alar plate are the target of sensory input to the spinal cord and brain. The Homeobox gene Nkx2.2 is expressed in the ventral part of the basal plate; Msx is expressed in the dorsal part of the alar plate. The expression of these two genes in a cross section of the zebrafish spinal cord are shown in (D). A third homeobox gene, Gsh-1, is expressed at a lateral level in between Msx and Nkx2.2. (E and F) Homologs of these genes are expressed in the same dorso-ventral succession in *Drosophila*. The schematic cross section of the fly neurectoderm (E) shows that the Nkx2.2 homolog Vnd forms a medial stripe adjacent to the midline cells; Ind (homolog of Gsh-1) and Msh (homolog of Msx) are expressed at an intermediate and lateral level, respectively. Panel F shows a ventral view of a whole mount *in situ* preparation demonstrating the expression of Vnd and Msh.

partitioned into neuromeres, each neuromere delineated by its own peripheral nerve. In the spinal cord of mammals, one distinguishes 7 cervical (neck) neuromeres whose neurons innervate the upper extremity, 12 thoracic neuromeres, 5 lumbar neuromeres belonging to the lower extremity, and a variable number of sacral neuromeres. Seven or eight neuromeres are present in the hindbrain, where they are called rhombomeres. Similar to spinal neuromeres, rhombomeres are characterized by the pattern of peripheral nerves they emit. The midbrain is thought to represent a single neuromere. The forebrain possesses no peripheral nerves; however, based upon other morphological landmarks, as well as the expression of molecular markers, it is subdivided into six prosomeres. The posterior three prosomeres (P4, P5, and P6) give rise to the diencephalon, while prosomeres 1, 2, and 3 form the telencephalon. The vertebrate neural tube is divided in the dorsoventral axis into four longitudinal domains, called the floor plate, basal plate, alar plate, and roof plate. The basal plate next to the floor plate includes all motor neurons. In the lateral wall of the neural tube, a furrow called sulcus limitans separates the basal plate ventrally from the alar plate dorsally. The alar plate later gives rise to the dorsal sensory column. Both the anterior-posterior and the dorsal-ventral axes of the neurectoderm are regulated by the expression of conserved homeodomain transcription factors (see Chapter 13). Thus, in both vertebrates and invertebrates, these spatially arrayed transcription factors of the anterior-posterior and the dorsal-ventral axes collaborate to divide the developing nervous system into grids of molecular coordinates that specify the positional identity of neural progenitors down to the level of small groups or, in some invertebrates, even single cells.

Neurogenesis is patterned in time as well as space. This is exemplified in *Drosophila* neuroblasts, which go through a program of proliferation where each neuroblast gives rise to a series of distinct GMCs. The first GMCs of a neuroblast lineage tend to lie deeper in the CNS and usually generate neurons with long axons, whereas the later arising GMCs stay closer to the edge of the CNS and produce neurons with short axons. In the stages when the first GMCs are generated, most neuroblasts express *Hunchback (Hb)*, and GMCs that arise at this time inherit this *Hb* expression. Later, the same neuroblasts turn off *Hb* and express *Krueppel (Kr)* instead, and GMCs generated at this stage inherit *Kr* expression. Experiments in which *Hb* or *Kr* is eliminated or expressed at the wrong time lead to predictable switches in the fates of early and later-born descendants of these neuroblasts (Isshiki, Pearson, Holbrook, & Doe, 2001). It is likely that a temporal succession of transcription factors like those in *Drosophila* also patterns neurogenesis in vertebrates (Elliott, Jolicoeur, Ramamurthy, & Cayouette, 2008).

In many CNS tissues there is a clear relationship between cell birth and cell fate. For example, during the development of the *Drosophila* nervous system, GMCs and immature neurons form a stack on top of the neuroblast from which they originated. Postmitotic neurons generally do not migrate, so the progeny of a neuroblast remain spatially close to one another, with the position of each neuron dependent on the position of the neuroblast and the time at which it was born. Thereby, a two-layered cortex-neuropil architecture typical of the mature ganglion is generated (Fig. 14.1H). The histogenesis of the six-layered, mammalian cerebral cortex is also easy to describe. It is "inside out" with the earliest cells to exit the cell cycle forming the deepest layers and the last cells to be born forming the most superficial layers. While intrinsic temporally controlled transcription factors play a key role in regulating histogenesis, it is also becoming clear that environmental factors can also influence histogenesis. For example, in the retina, the first cells to be born are retinal ganglion cells, which by negative feedback influence the next born cells to choose a different fate—by downregulating the transcription factors that are important for the determination of retinal ganglion cells. There is also a rich dialogue between factors that drive cell cycle exit and cell fate, such that some transcription factors (see below) have direct effects on the cell cycle, and some cell cycle exit factors influence cell fate (Ohnuma & Harris, 2003).

THE PRONEURAL AND NEUROGENIC GENES

In the developing nervous system of *Drosophila*, the functions of the genes known as the proneural and the neurogenic genes, which have proved to be so important for the initial differentiation of neurons in all animals, were first elucidated. The neurectoderm of the *Drosophila* embryo is a mixed population of cells. Only some of these cells become neuroblasts; others develop as dermoblasts. Cell–cell interactions that take place within the neurectoderm define the number and pattern of neuroblasts. Experimental and genetic studies suggest a two-step mechanism for this process (Campos-Ortega, 1995). First, discrete groups of neurectodermal cells are made competent to become neuroblasts. These groups of cells, called proneural clusters, represent "equivalence groups" in which all cells initiate a neural rather than an epidermal fate (Fig. 14.3A, B). In *Drosophila*, a group of proneural genes expressed in the proneural clusters is responsible for making these cells competent to become neural progenitors. In a second step, cells of the proneural cluster interact to sort out which of them will become neurons and which will fall back to become dermoblasts (Fig. 14.3A). As the first cell in each cluster begins to differentiate as a neuroblast, it sends out inhibitory signals to its neighbors, inhibiting them from also becoming neuroblasts. This can be demonstrated by ablating a developing neuroblast with a laser micro-beam,

as what happens next is that a neighboring cell of the equivalence group, which would otherwise have developed into dermoblast, switches its developmental program and replaces the lost neuroblast. It is the neurogenic genes that mediate this inhibitory cell–cell interaction.

Most proneural genes are transcription factors of the basic helix-loop-helix (bHLH) family. Loss of proneural gene function leads to a reduction in neuron production. Misexpression of proneural genes often has the opposite phenotype. For example, misexpression of proneural gene mRNA in vertebrates leads to an increased number of neural progenitor cells in the neural tube or transforms of epidermal cells into neurons (Fig. 14.3D, E). The neurogenic genes encode a cell communication mechanism that

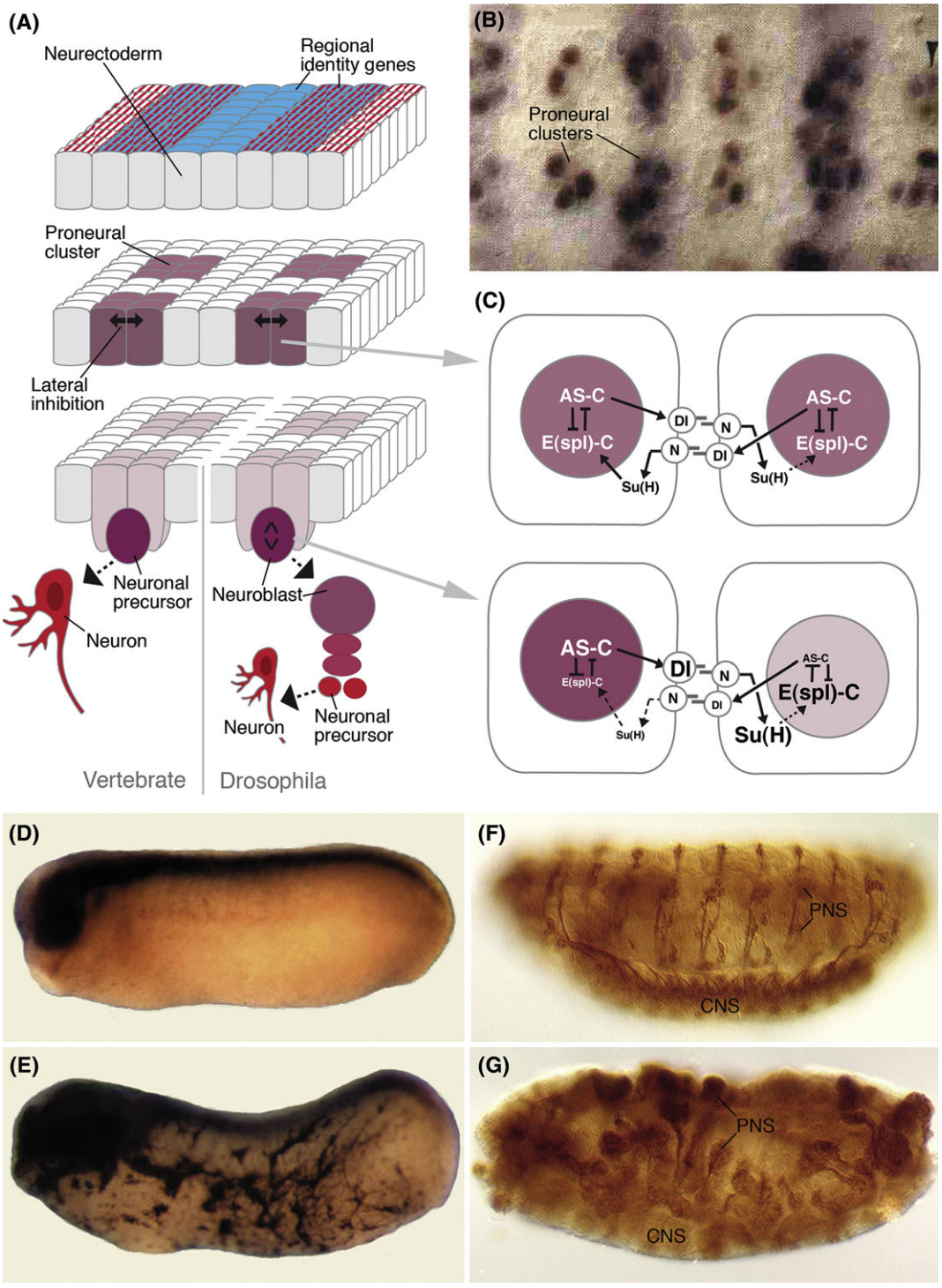


FIGURE 14.3 Function of proneural and neurogenic genes in neural development. (A) The sequence of events that leads to the segregation of individual neural progenitors (in insects) or neurons (in vertebrates). First, proneural genes are turned on in discrete cell populations, called proneural clusters (purple). In *Drosophila*, a number of genes (called "regional identity genes" in the diagram) have been identified. They are expressed in a distinct pattern of transverse and longitudinal stripes before proneural genes appear and are proven or likely factors to control the expression of proneural genes. Proneural genes in turn trigger different events in the proneural clusters. One outcome of their expression is a cell–cell interaction process, mediated by the neurogenic genes, that selects a single cell (or a smaller group of cells) from each proneural cluster. Only this cell continues along the neural pathway and segregates from the neurectoderm as a neuroblast (right side of panel; intense purple), which divides and produces a neural lineage (red); the other cells of the proneural cluster remain within the neurectoderm. In vertebrates (left side of panel), the entire neurectoderm (i.e., the neural plate/neural tube) is formed by neural progenitors. Here, the proneural/neurogenic gene cassette seems to select cells that become postmitotic and differentiate from other (neural progenitor) cells, which remain in the neuroepithelium and continue to proliferate. (B) Detail of proneural gene expression in the neurectoderm of the *Drosophila* embryo (*from Skeath & Carroll, 1992*). (C) Interaction between proneural and neurogenic genes in the *Drosophila* neurectoderm. The upper panel shows two cells of a proneural cluster in which proneural genes of the *achaete-scute* complex (AS-C) are expressed. Genes of the AS-C activate the expression of the signaling molecule Delta (Dl). Dl activates the receptor Notch (N) in neighboring cells; this leads to cleavage of the N molecule and the translocation of the cytoplasmic domain of N along with the transcription factor Suppressor of Hairless [Su(H)] to the nucleus, where they upregulate the expression of proteins of the Enhancer of split complex [E(spl)-C]. E(spl) initiates or maintains the development of an undifferentiated neurectodermal cell; at the same time, it directly inhibits the expression of AS-C genes, which promote differentiation of a cell as a neural progenitor or neuron. Imbalances introduced by an unknown mechanism into the levels of AS-C or E(spl)-C expression, respectively, are rapidly amplified (lower half of C) and lead to the segregation of a single neural progenitor (high levels of AS-C) from other cells that stay epithelial [high levels of E(spl)-C]. (D–G) The proneural gene turns epidermal cells into neurons. (D) A normal *Xenopus* embryo stained for the neural marker NCAM shows no staining in the epidermis. (E) A neuroD-injected embryo has NCAM-stained cells with neuronal morphology in the entire epidermis. (F) Wild-type *Drosophila* embryo labeled with the neuronal marker 22C10 contains discrete clusters of neurons forming the central nervous system (CNS) and peripheral nervous system (PNS). (G) Loss of the signaling molecule Dl causes a massive increase in neurons at the expense of epidermis.

acts to restrict the number of developing neurons. The neurogenic genes are also known as the Notch signaling pathway. The *Delta* and *Notch* genes encode membrane proteins that function as signal and receptor, respectively. Activation of Notch by Delta sets in motion a signal transduction cascade (Fig. 14.3C) that involves another neurogenic gene called *Suppressor of Hairless* in *Drosophila* or CBF-1 in mammals. Su(H)/CBF-1 forms a complex with the intracellular domain of Notch, which is released from the membrane when the pathway is activated. This complex moves into the nucleus and activates the *E(spl)-C* genes in *Drosophila* or the ESR and Hairy genes in mammals. These are bHLH transcription factors that antagonize and repress the proneural bHLH genes, completing an inhibitory feedback loop. In *Drosophila* loss of function of any of the neurogenic genes results in a higher number of neural progenitors, at the expense of dermoblasts (Fig. 14.3F, G). In mammals, loss of the Notch signaling pathway usually results in a higher number of a specific type of neuron at the expense of other types or at the expense of glial cells, depending on when and where this function is lost.

TRANSCRIPTIONAL HIERARCHIES AND NETWORKS

As we have seen above with respect to spatial and temporal determination as well as in the proneural and neurogenic pathways, many of the key players in neuronal determination are transcription factors. Transcription factors are protein products of genes that turn on or repress the transcription of other genes or themselves. We find again and again, whether in simple worms or mammals, that transcription factors have a central role in cell-fate determination. Often we find complex transcriptional hierarchies whereby one transcription factor induces the expression of another, which induces the expression of yet another. Such linear hierarchies often make fate decisions progressive and incremental. Such a drive forward in the cell fate progression is sometimes accomplished by a more complex network of transcription factors such that each one activates the next one and represses the previous one so that once a progenitor turns on a particular set of intrinsic factors, it can no longer revert to a more primitive state. Once a particular state is reached, transcription factors can work in interactive circuits that reinforce decisions by mutual cross-activation. But, as we shall also see, some networks of transcription factors work by cross-repression to sharpen boundaries between domains and to make sure that neighboring cells have distinct fates. We also note that a single progenitor will express a variety of transcription factors at any one time and that these can often act together as a combinatorial transcription factor code. Theoretically, a set of just 4 transcription factors on or off in a cell could specify one of 16 distinct fates. The developing nervous system expresses hundreds or even thousands of different transcription factors, so it isn't difficult to imagine how transcriptional codes could finely specify the fate of billions of distinct neuronal types.

ASYMMETRIC CELL DIVISION AND CELL FATE

Neural progenitors, as we mentioned above, can divide symmetrically to increase geometrically in number; or they divide asymmetrically, whereby one daughter cell

retains progenitor fate and keeps dividing, whereas the other one differentiates as a neuron. To produce these two distinct fates, intrinsic determinants expressed in the neural progenitors are often differentially distributed during asymmetric divisions. The mechanism controlling asymmetric divisions has been elucidated in great detail for the *Drosophila* neuroblasts, which, as we know, generate one large daughter neuroblast that will divide again and a small GMC that divides just once more to produce postmitotic neurons.

The machinery involved in asymmetric cell division (Betschinger & Knoblich, 2004) involves a number of proteins at the apical and basal cortex of the neuroblast (Fig. 14.4A). At the apical side of the neuroblast is a highly conserved complex of localized proteins called the "Par-complex." Members of this complex include the so-called Par proteins (Par3, Par-6) and atypical protein kinase C (aPKC). As the neuroblast delaminates from the neurectoderm, the Par complex maintains its polarized localization, defining the apical pole of the cell (Fig. 14.4A).

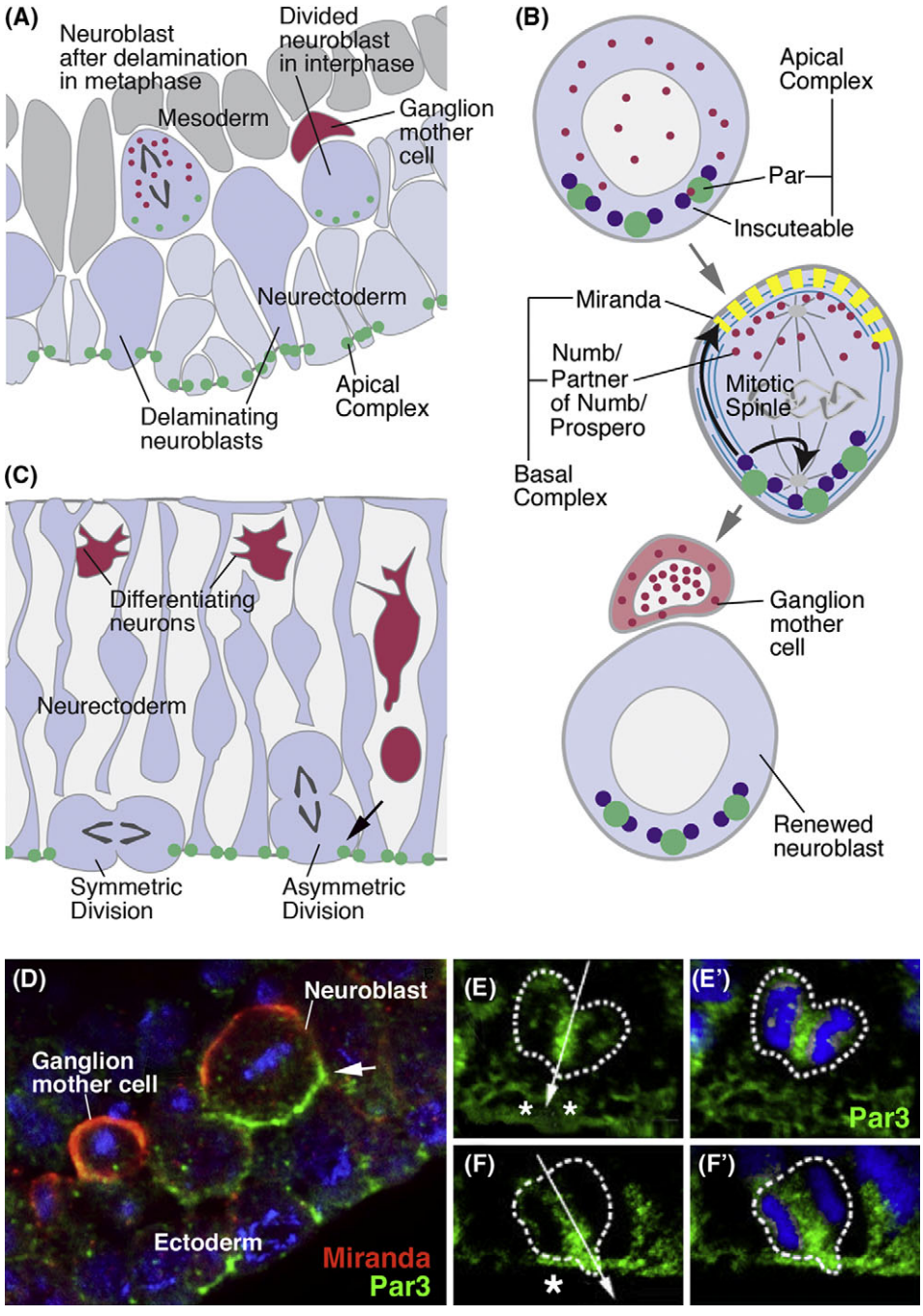


FIGURE 14.4 Control of asymmetric cell division in neural progenitors. (A) Schematic section of *Drosophila* neurectoderm at the stage of neuroblast delamination. Proteins of the apical complex (e.g., Par3, Inscuteable) are expressed apically in the neurectoderm and are carried interiorly by delaminating neuroblasts. (B) Schematic of neuroblast before (top), during (middle), and after (bottom) mitosis. The apical complex controls asymmetric distribution of intrinsic fate determinants, such as Numb and Prospero (red), by orienting the mitotic spindle vertically and by localizing the Miranda protein (yellow) basally. Miranda "traps" Numb at the basal pole of the dividing neuroblast (middle panel) and thereby channels it into the ganglion mother cell (bottom panel). (C) Schematic cross section of the mouse neuroepithelium. The mouse Par complex is also expressed apically in neural progenitors (green). In asymmetric divisions, Par 3 is maintained only in the epithelial daughter cell and not the delaminating neural precursor (arrow). (D) Confocal section of *Drosophila* neurectoderm (*from Wodarz & Huttner, 2003*). Expression of apical complex (represented by Par3, green) in ectoderm (arrowhead) and delaminated neuroblast (arrow). Miranda (red) is expressed basally; following neuroblast mitosis, it is restricted to ganglion mother cell. (E–F') Confocal sections of dividing neural progenitors in mouse neural tube (neurectoderm), showing expression of Par3 (*from Bultje et al., 2009*). (E, E') Symmetric division. Apical expression of Par3 (small asterisks) is inherited by both daughter cells. (F, F') Asymmetric division. Par3 ends up only in the one daughter cell to the left; the other daughter cell, lacking Par3, delaminates and adopts a neural fate.

During neuroblast divisions, the Par-complex recruits another group of proteins, Inscuteable (Insc) and Partner of Inscuteable (aka Pins), to the apical pole (Fig. 14.4B). The importance of the apically localized Par/Insc complex for asymmetric cell division is twofold: (1) Pins locally activates a G-protein-coupled pathway that attracts one of the centrioles toward this complex. The result of this is that the axis of the mitotic spindle is apicobasal (Fig. 14.4B). An apicobasally directed spindle, resulting in an apical daughter cell and a basal daughter cell, constitutes a general feature of asymmetric divisions in the developing nervous system; by contrast, symmetric divisions are generally characterized by horizontal spindles. (2) Proteins of the Par-complex interact with the cytoskeleton in such a way that a group of other proteins, which are also highly conserved among different animal groups, are transported toward the basal pole of the cell. These basal proteins include Miranda (Mir), Numb, and Prospero (Fig. 14.4B, D). As a result of the apicobasal mitotic spindle, only the basal daughter (i.e., the GMC) cell will receive these proteins. Prospero is a homeodomain transcription factor whose function is to switch on genes required for neuronal differentiation and switch off genes that are crucial for continued proliferation (Choksi et al., 2006). Numb acts as a negative regulator of the Notch pathway (discussed above), which keeps Notch pathway activity low in the GMC. Low Notch activity causes the de-repression of proneural genes, as well as the arrest of proliferation (after one more mitoses) of the GMC. By contrast, in the basal daughter cell, Notch pathway activity is kept high by proteins of the apical complex (Par-3, Insc), which promotes continued division of this cell.

In the vertebrate CNS, neuroepithelial cells first tend to divide symmetrically, giving rise to two dividing progenitor cells, but as neurogenesis proceeds, some neuroepithelial cells begin to divide asymmetrically—in this case giving rise to one neuroepithelial progenitor and one differentiated neuron. The mechanism responsible for vertebrate asymmetric divisions, including many of the genes involved, is similar to what has been described above for *Drosophila*.

In vertebrates, the orientation of the spindle axis is also thought to play a decisive role in switching between symmetric and asymmetric divisions. Divisions along the horizontal plane produce two cells that stay in contact with the ventricular (apical) surface, whereas more apicobasal divisions produce a basal daughter that migrates away from the ventricular surface, while the other daughter remains in contact, consistent with it retaining a neuroepithelial fate (Fig. 14.4C) (Chenn & McConnell, 1995). Recent findings show that mitotic spindles that are only slightly tilted away from the horizontal axis are sufficient to generate daughters with different fates. Neuroepithelial cells are highly columnar and have only a small apical membrane domain contacting the ventricular surface. As in *Drosophila*, the Par-complex proteins are restricted to the apical domain. What appears to matter in determining the fate of a daughter cell is whether or not it inherits the apical domain (Wodarz & Huttner, 2003). Only perfectly apicobasal cleavage planes will cut through the apical domain, such that both daughters will receive half of it. In this case, the mitosis is symmetric, producing two progenitors that remain epithelial and continue to divide. If the plane is tilted, even if only slightly, one daughter cell will inherit all of the apical domain and remain proliferative; the other one will delaminate and differentiate (Fig. 14.4.E, F).

Among the vertebrate homologues of the *Drosophila* genes involved in asymmetric cell divisions are several members of the apical complex. Par-3 acts as an activator of Notch pathway activity in the neuroepithelium, maintaining progenitor proliferation. mInsc is the mammalian homologue of Insc, and ASG3 is the mammalian homolog of Pins. Interfering with the expression or function of these homologues leads to the loss of apicobasal spindle orientation in the mammalian nervous system, resulting in too many cells dividing in the horizontal plane. In the mouse retina this causes the production of large clones arising from neuroepithelial cells, which can be explained if these cells divide symmetrically to produce two progenitors rather than one. Several members of the basal complex, such as Numb and Prospero, may also play a key role in the distinct fates of daughter cells, for example, knockout of the mammalian homolog of Numb is essential for cortical neuroepithelial cells to remain in a progenitor state (Petersen, Zou, Hwang, Jan, & Zhong, 2002).

The orientation of division may thus have an important role in determining cell fate by the uneven segregation of determinants between the daughter cells, but it can also cause the two daughter cells to inhabit slightly different chemical niches—for example, one daughter might remain closer to a source of a secreted growth factor or an extracellular differentiation factor.

NEURONS AND GLIA

The nervous system is composed of two main classes of cells: neurons and glia. Glial cells of the vertebrate CNS include three major types, microglia, astrocytes, and oligodendrocytes. Microglia are derived from the bone marrow and will not be further considered here. Astrocytes are multipolar cells whose processes interact with neuronal synapses, as well as the capillary network; astrocytes form the blood-brain barrier. Oligodendrocytes produce large, lamellar processes that make up the myelin sheath around axons.

Lineage tracing experiments show that neurons and glial cells are often produced from a common progenitor, which we can call a "neuroglioblast" in *Drosophila*. Neuroglioblasts form part of the neuroblast population that arises

in the embryo (Fig. 14.5A, B). Two generic determinants of glial fate have been identified. These include, *glial cells missing (gcm)* and *reversed polarity (repo)*, both of them expressed at an early stage in neuroglioblast development (Jones, 2005). Their crucial involvement in glial fate is attested to by the fact that ectopic expression of these genes converts neurons into glial cells and that loss-of-function mutants lack glial cells. Interestingly, neither *repo* nor *gcm* appear to play a role in glial development in vertebrates.

In the central nervous system of vertebrates, as in *Drosophila*, multipotent neural progenitors give rise to glia, as well as neurons (Fig. 14.5C). It is usually the case that neurons arise first and glia later. The first rounds of division of the neural progenitors are symmetric and lead to an increase of the progenitor pool (Fig. 14.5D1). Subsequently, asymmetric divisions give rise to neurons (Fig. 14.5D2). Neurons move away from the ventricular surface and form a mantle layer of increasing thickness.

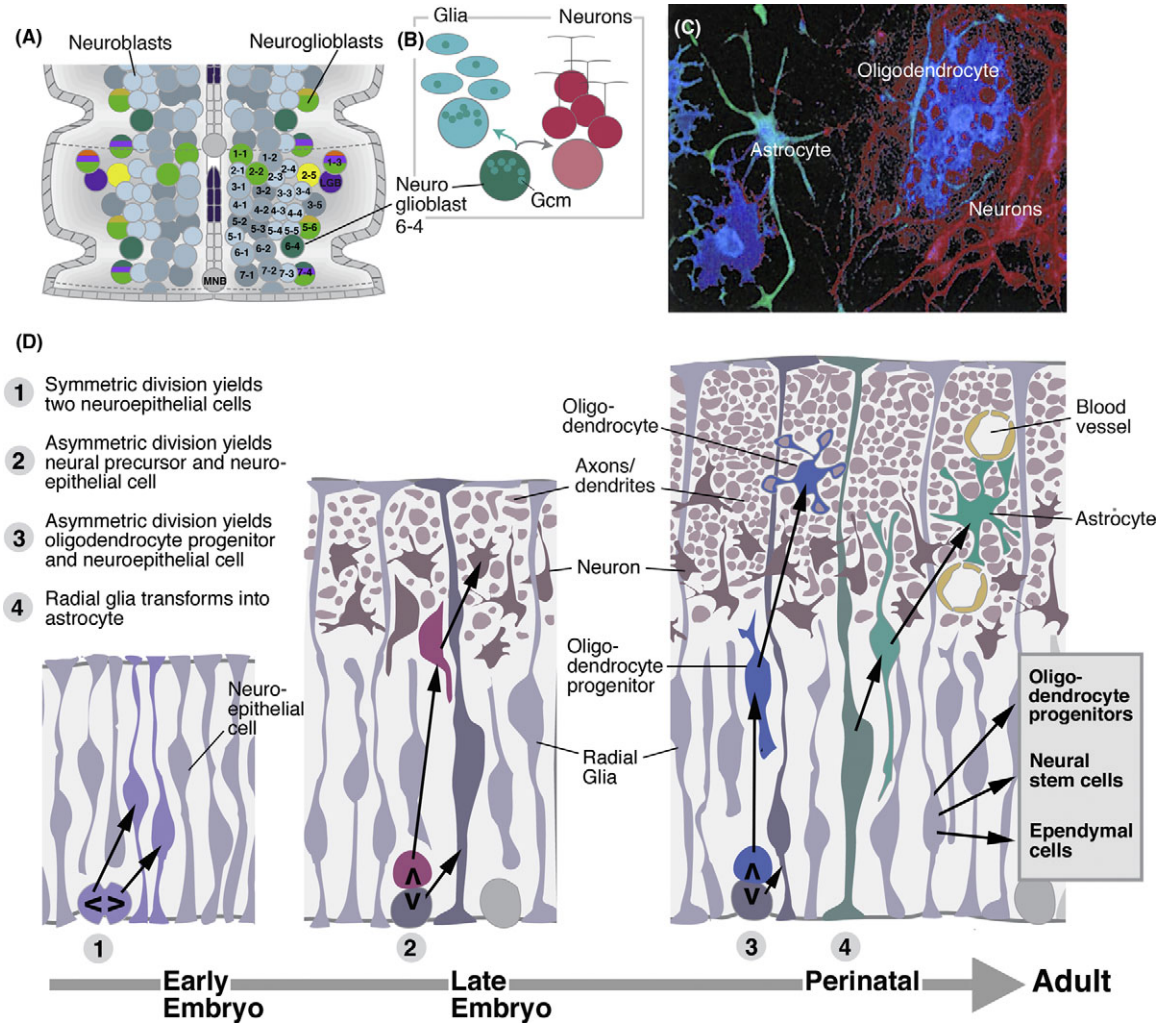


FIGURE 14.5 Specification of glial lineages. (A) Neuroblast map of the ventral nerve cord with glioblasts and neuroglioblasts indicated. Neuroblasts/neuroglioblasts of one hemineuromere are identified alphanumerically. (B) Neuroglioblast Nb6-4 expresses the glial regulatory protein Gmc (green). When this cell divides into two daughter cells, 6-4 G and 6-4 N, the Inscuteable complex and Miranda segregate Gmc into 6-4 G, which thereby becomes specified as glioblast. 6-4 N generates neurons. (C) Photograph of cell lineage obtained by injecting an individual neuroepithelial cell from mouse cortical progenitor cultured in a dish. Neural and glial-specific antibodies reveal the presence of both neurons (purple) and glia (astrocytes: green; oligodendrocytes: blue) in the lineage (*from Qian et al., 2000*). (D) Schematic cross section of neural tube of early embryo (left), late embryo (middle), and around birth (right; see time line at bottom). At early stage, symmetric divisions of neuroepithelial cells (1) result in an increased number of neural progenitors. At later stages, neuroepithelial cells (now synonymous with radial glial cells) start to divide asymmetrically, producing neurons (red), and maintaining their own number. Perinatally, production of neurons ceases, and neuroepithelial cells/radial glia switch to the production of oligodendrocyte progenitors (blue). Many radial glial cells delaminate and become astrocytes (green). Radial glia also give rise to the ependymal cells lining the ventricles of the adult CNS, to more oligodendrocyte progenitors, and to neural stem cells that remain active throughout adult life (*after Kriegstein & Alvarez-Buylla, 2009*).

As a result, toward the end of neurogenesis, the neuroepithelial cells that are left become elongated cells, called radial glia (Fig. 14.5D2). These cells still have their original apicobasal polarity; with their apical poles they form the lining of the ventricle; the long, basal process (that serves migrating neurons as guidance cue; see Chapter 15) is in touch with the basal surface, where blood vessels and meninges (the connective tissue layers surrounding the CNS) appear. Radial glia keep dividing, now mostly producing oligodendroctye progenitors (OLPs; Fig. 14.5D3). OLPs migrate throughout the central nervous system and differentiate into oligodendrocytes; many OLPs remain mitotically active throughout adult life. At the same time, most of the radial glia withdraw their apical attachments to the ventricular layer and differentiate into multipolar, mature astrocytes (Fig. 14.5D4). Astrocytes continue to divide throughout life; they react to neural injury by strong proliferation, forming glial scars (gliosis). A few radial glia, at specific locations, form a reservoir of adult neural stem cells (see Box 14.1).

BOX 14.1

NEURAL STEM CELLS

The proliferating progenitor cells or neuroblasts that give rise to more differentiated progeny but themselves remain in the cell cycle are called neural stem cells (NSCs) (Gage, 2000). Some frogs and fish continue to grow bigger throughout their lifetimes, and their brains grow with their bodies, adding new cells from localized stem cell populations. However, in most animals, including humans, the majority of proliferating neural cells use themselves up during development so the sources of new neurons in an adult animal are extremely limited. This is why damage to the central nervous system is medically much more serious than damage to the other organs, such as the skin or liver, where stem cells that persist into adulthood can replace injured tissue. Nevertheless, it has been found that even in adult mammals, NSCs persist in certain "niches" near ventricular layers. Interestingly, one pocket of rich stem cell activity is in the hippocampus, where certain aspects of learning take place. It is thought that a turnover of pyramidal cells in the hippocampus may be important for replacing old memory traces with new ones.

There is excitement about the potential of using NSCs in a replacement strategy for brain damage due to injury, stroke, or degenerative diseases such as Parkinson's disease, or retinal degeneration. One could imagine using NSCs to treat patients suffering from Parkinson's disease or Retinitis Pigmentosa. The cells could be expanded in culture under conditions where they begin to differentiate along particular pathways such as dopaminergic neurons or photoreceptors, and these new neurons could be transplanted into the brain or retina to replace ones lost due to damage or disease.

Important questions about neural stem cells in the adult brain abound. Why do these cells remain undifferentiated and capable of division when their neighbors have exited the cell cycle and differentiated? What signals do these cells need in order to stimulate their proliferation and differentiation? How can their differentiation toward particular types of neurons or glia be controlled experimentally? While there is much speculation on the role of various signaling and growth factors in regulating these processes, more work needs to be done to identify the molecules that prevent stem cells from differentiating and the signals that release their potential.

Interestingly, the environment of the organism plays a key role in regulating NSCs (Kempermann, 2011). Exercise is clearly important for the generation of new neurons in the adult rat hippocampus, and an enriched environment is important for the survival of new generated hippocampal cells. Rats put in simple cages without exercise wheels generate fewer new neurons, and these neurons have a higher chance of dying than in rats that are given exercise regimes and who are held in complex cages with a variety of toys. Stress works in the other direction and seems to inhibit stem cell proliferation and survival in the hippocampus. The identification of the regulatory factors that regulate these NSC activities might prove invaluable in treating neural or glial degenerative conditions without the need for transplants.

Another source of stem cells in mammals is the inner cell mass of the early conceptus. These are the cells that give rise to all the tissues of the embryo proper (Vallier & Pedersen, 2005). When these embryonic stem cells (ESCs) are grown in culture under certain conditions, they can produce neural precursors. It seems that the lessons learned about the mechanisms of neural induction are paying off in this context, because neural inducing signals increase the probability that ESCs will follow a neural pathway. ESCs may have an advantage over neural stem cells from adults because they come from a stage in development when their potential fates are less restricted by the inheritance of intrinsic determinants. More recently, using a cocktail of transcription factors that are normally expressed in ESCs, it has become possible to turn almost any cell in the body back to a pluripotent stem cell (Takahashi & Yamanaka, 2006).

BOX 14.1 (*cont'd*)

Stem cells produced in this way are called induced Pluripotent Stem Cells (iPSCs) and can be taken from the skin or blood of patients. Human iPSCs, besides having a potential role in therapy for neuronal injury and disease, are also very useful for the cellular modeling of disease processes, such as certain forms of autism (Marchetto et al., 2010).

Various laboratories are trying to turn ESCs or iPSCs into various types of neurons. Since a great deal is known about the inductive signals that are critical to the generation of motor neurons in the vertebrate spinal cord, it may be asked whether this knowledge can be used to direct stem cells to a motor neuron fate. Jessell and colleagues (Wichterle et al., 2002) combined inducers that first turn mouse ESCs into neural progenitors with factors such as retinoic acid that drives posterior regional identity such as spinal cord and Shh that drives ventral fates in the spinal cord to push these ESCs to differentiate into neural then spinal and finally into motor neurons that express the same combinatorial codes of transcription factors as motor neurons *in vivo*. These ESC-derived motor neurons can be labeled by Green Fluorescent Protein (GFP) under the control of the HB9 promoter (expressed only in motor neurons) and when these fluorescents cells are transplanted into the embryonic spinal cord of a chick embryo, they extend axons, and form synapses with target muscles (Fig. B14.1). Moreover such ES-derived motor neurons develop appropriate transmitter enzymes and receptors, and electrical properties, such that they can make perfectly functional synapses with muscle fibers. Thus, by using signaling pathways discovered by studying normal cell determination in the developing nervous system, biologists may be able to direct stem cells to form specific types of neurons.

Clearly, the more we know about neuronal determination, the more likely we will be able to direct stem cells down appropriate developmental pathways that will be useful in treating the damaged nervous system.

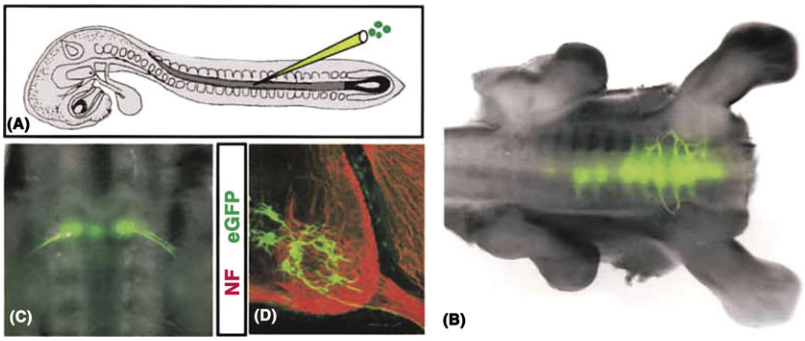


FIGURE B14.1 Integration of transplanted ESC-derived motor neurons into the spinal cord. (A) Implantation of fluorescently labeled ESC-derived motor neurons into an embryonic chick spinal cord. (B) Fluorescence image showing GFP-labeled motor neurons in thoracic and lumbar spinal cord. (C and D) Location of ES-cell derived GFP-labeled motor neurons clustered in the ventral spinal cord and sending axons out the ventral root. (*From Wichterle et al., 2002.*)

Oligodendrocytes arise during several waves. During the first wave, high levels of the ventrally expressed Shh morphogen activate transcription factor determinants of oligodendrocyte fate, called Olig 1 and Olig 2, in the ventral spinal cord (Kessaris, Pringle, & Richardson, 2001; Zhou, Choi, & Anderson, 2001). At later developmental times, oligodendrocyte and astrocyte formation occurs at more dorsal levels of the neural tube. Early expressed transcriptional regulators akin to Olig have not yet been identified for astrocytes. However, we do know that BMPs and some cytokines (e.g., ciliary neurotrophic factor, CNTF) promote astrocyte fate. BMPs and Shh act in an antagonistic manner in the decision between astrocyte and oligodendrocyte formation: BMP activates astrocyte formation and inhibits oligodendrocytes, while Shh does the reverse.

The Notch signaling pathway also plays an important role in glial development. First, high Notch activity in the neuroepithelium acts as a permissive factor for gliogenesis. Thus, activation of the Notch pathway suppresses the transcription of proneural bHLH transcription factors. Lowering Notch levels will result in premature conversion of neuroepithelium into neurons, thereby depleting the pool of cells that would normally give rise to glia at a later stage. Zebrafish embryos with reduced Notch activity in the spinal cord show increased numbers of neurons, at the expense of glia. In addition, at a later stage, Notch may also play a direct role in promoting the expression of glial specific genes. Introducing a constitutively active form of Notch into cultured neural progenitors triggers astrocyte differentiation.

SENSORY NEURONS OF THE PERIPHERAL NERVOUS SYSTEM

With the exception of the eye, all sensory organs of the vertebrate are derived from sensory placodes and neural crest. A placode is a morphologically visible thickening of an epithelium that often subsequently invaginates into the interior and transforms into a distinct tissue or organ (Schlosser, 2006). Sensory placodes appear during the late stages of neurulation in the dorsal ectoderm that flanks the invaginating neural tube (Fig. 14.6A–C). The most anterior placode is the olfactory placode that gives rise to the olfactory sensory neurons of the nose. Further posterior is the otic placode that forms the inner ear. Slightly more laterally is the placode of the trigeminal system that produces the sensory neurons innervating the receptors of the skin and muscles of the face and the head. Developing later and at a more lateral level in contact with the pharyngeal clefts (the gill slits in fish and amphibians) are the epibranchial placodes. They give rise to sensory neurons that innervate the taste receptors of the mouth, and receptors in the inner organs, such as the heart, large blood vessels, and digestive tract.

Sensory neurons of the trunk are derived from the neural crest (Fig. 14.6B–D). The neural crest is a transient population of cells that arises along the lateral edges of the neural plate (Chapter 13). Crest cells become localized to the dorsal part of the neural tube as it folds up, and then leave the neural tube and migrate along several well-defined pathways. Cells destined as sensory neurons aggregate in segmentally reiterated masses that form the dorsal root ganglia. Long peripheral dendrites extend from the dorsal root ganglia toward the skin, muscles, and skeleton. Here, dendritic endings of sensory neurons interact with specialized, mesodermally derived support cells, forming small "corpuscules" that are tuned to a specific type of stimulus, including touch, vibration, stretch, temperature, and pain.

Classical embryological experiments in which parts of donor embryos were transplanted to different sites in host embryos indicated that as they invaginate, placodes come in contact with the neural tube, the archenteron, the mesoderm, and the epidermal ectoderm. These structures all have inductive capacities and influence the fate of the placodal cells. In some instances, specific signaling molecules have been identified. The picture that emerges is that neural crest and sensory placodes are specified in a stepwise manner from the border territory in between neural plate and epidermal ectoderm. (1) At an early stage, prior to neurulation, BMP signals expressed in the epidermal ectoderm and FGF from mesoderm underlying the ectoderm trigger the expression of "border genes," including the homeobox transcription factors Msx1 and Dlx3/5 (Fig. 14.6A). (2) As neurulation sets in, the border territory splits into neural crest, medially, and preplacodal ectoderm, laterally (Fig. 14.6B). BMP expression reaches peak levels in, and is required for, neural crest ectoderm; at the same time, BMPs inhibit placode formation. Coinciding with high BMP activity are Wnt signals. Both signaling pathways promote the expression of a new set of transcription factors required for neural crest development, including Sox10 and FoxD3. (3) Antagonists of BMP and Wnt are secreted from the mesoderm underlying the lateral part of the former border territory. In response, cells in this region turn on the transcription factors Six1/4 and Eya1/2, which specify the lateral border territory as preplacodal ectoderm (Fig. 14.6B). (4) The preplacodal ectoderm splits into the different sensory placodes as a result of signals from adjacent neural tube, mesoderm, and ectoderm. For example, FGF3 and Wnt8c from the hindbrain and FGF19 from the axial mesoderm work together to induce the otic placode (Fig. 14.6C, D). Each placode acquires its own characteristic set of transcription factors. Different members of the Six and Eya family, initially expressed in an overlapping pattern in the entire preplacodal ectoderm, become restricted to individual placodes. Another class of transcription factors, the Pax proteins, is added as well. Notably, *pax2* expression is strongly correlated with the otic placode; *pax6* appears in the olfactory and lens placode, and *pax3* and *pax8* in the trigeminal placode (Fig. 14.6C). Absence or downregulation of these genes results in loss of the corresponding sensory organ.

Homologues of most of the molecules involved in specification of sensory placodes also operate in the *Drosophila* sensory system. Indeed, molecular evidence shows that the proteins encoded by the *pax*, *eya*, and *six* genes form complexes that as a whole unit bind to DNA and regulate transcription. Genes that in this manner encode cooperating proteins are called gene networks. The fact that their function is dependent on all members of the complex explains why entire gene networks are often conserved throughout evolution. The *pax/six/eya* network, activated by BMP and/or Wnt signals, is one of the best-studied examples of such a conserved gene network.

Specification of the sensory placode is just the first step in a long chain of events leading up to the mature sensory organ. Let us take a look at this process for the otic placode that gives rise to all cells of the inner ear. Following invagination from the dorsal ectoderm, the otic placode forms a hollow vesicle, the otic vesicle or otocyst (Fig. 14.6E, F). The lumen of the otocyst ultimately becomes the endolymph filled cavity of the inner ear. Initially, the wall of the otocyst is formed by a homogenous epithelial layer. Subsequently, the otocyst folds in a complex manner to give rise to the cochlea for sensing sound, the utricle for sensing gravity, and the semicircular canals for sensing rotational head movements in the three dimensions (Fig. 14.6G). Again, signals acting upon the otic placode from the outside induce these different structures. For example, Shh from the floorplate of the neighboring neural tube acts on the ventral otocyst and specifies this territory as cochlea; ectopic

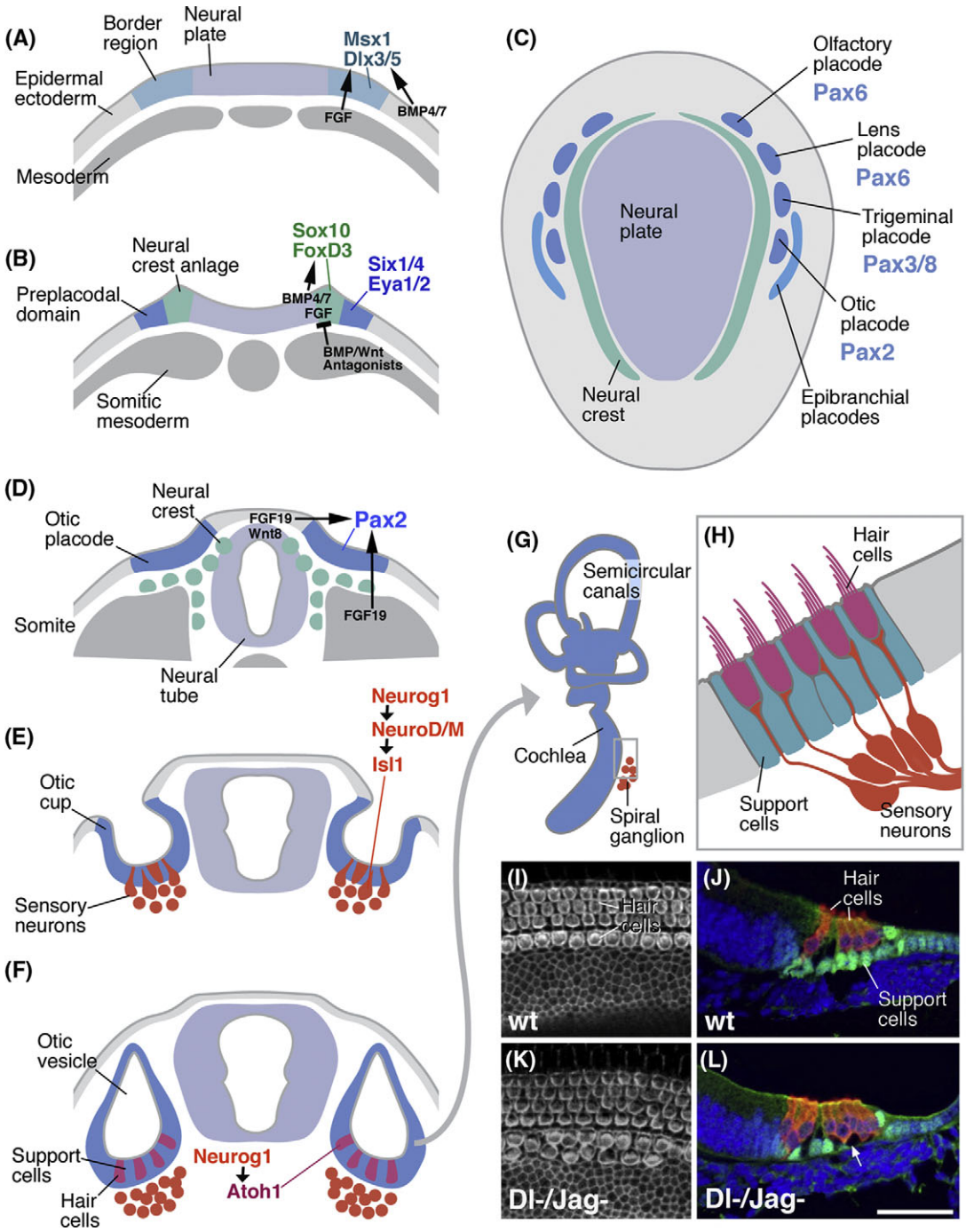


FIGURE 14.6 Sensory organ development in vertebrates. (C) Schematic dorsal view of neurula stage embryo, showing neural plate (light blue) flanked by sensory placodes (dark blue; anteriorly) and neural crest (green). Placodes express specific transcription factors of the Pax family. (A, B, D–F) Schematic cross sections of vertebrate embryo at successive stages, showing specification of the neural crest and sensory placodes (A, B) and development of the otic placode (D–F). (A) BMP signals from the lateral (epidermal) ectoderm and FGF signals from the mesoderm specify border region in between neural plate and epidermal ectoderm. (B) Peak levels of BMP and Wnt at the lateral boundary of the neural plate specify neural crest. Further laterally, these signaling pathways are inhibited, allowing for the appearance of the preplacodal ectoderm, from which all sensory placodes are derived. (D) Otic placode is induced from dorsal ectoderm by inductive signals (e.g., FDG19, Wnt8c) from neural tube and somites. (E) Otic placode invaginates. Precursors of sensory neurons (red) delaminate from placodal epithelium. (F) Cells of the placodal epithelium differentiate. Precursors of hair cells (sensory receptors; magenta) and support cells (green) appear. (G) Otic placode grows and forms cochlea and semicircular canals. (H) Magnified view of cross section of part of cochlea. Dendrites of sensory neurons (spiral ganglion) form synaptic contacts with sensory hair cells. These specialized, ciliated cells are surrounded by support cells. (I and J) Surface view (I) and section (J) of mouse cochlea, showing normal number and pattern of hair cells and support cells. (K and L) In Dl/Jag double mutants, hair cells are increased in number, and support cells are decreased. (*From Kiernan, Cordes, Kopan, Gossler, & Gridley, 2005.*)

expression of Shh in the entire otocyst causes lack of semicircular canals, and expansion of the cochlea.

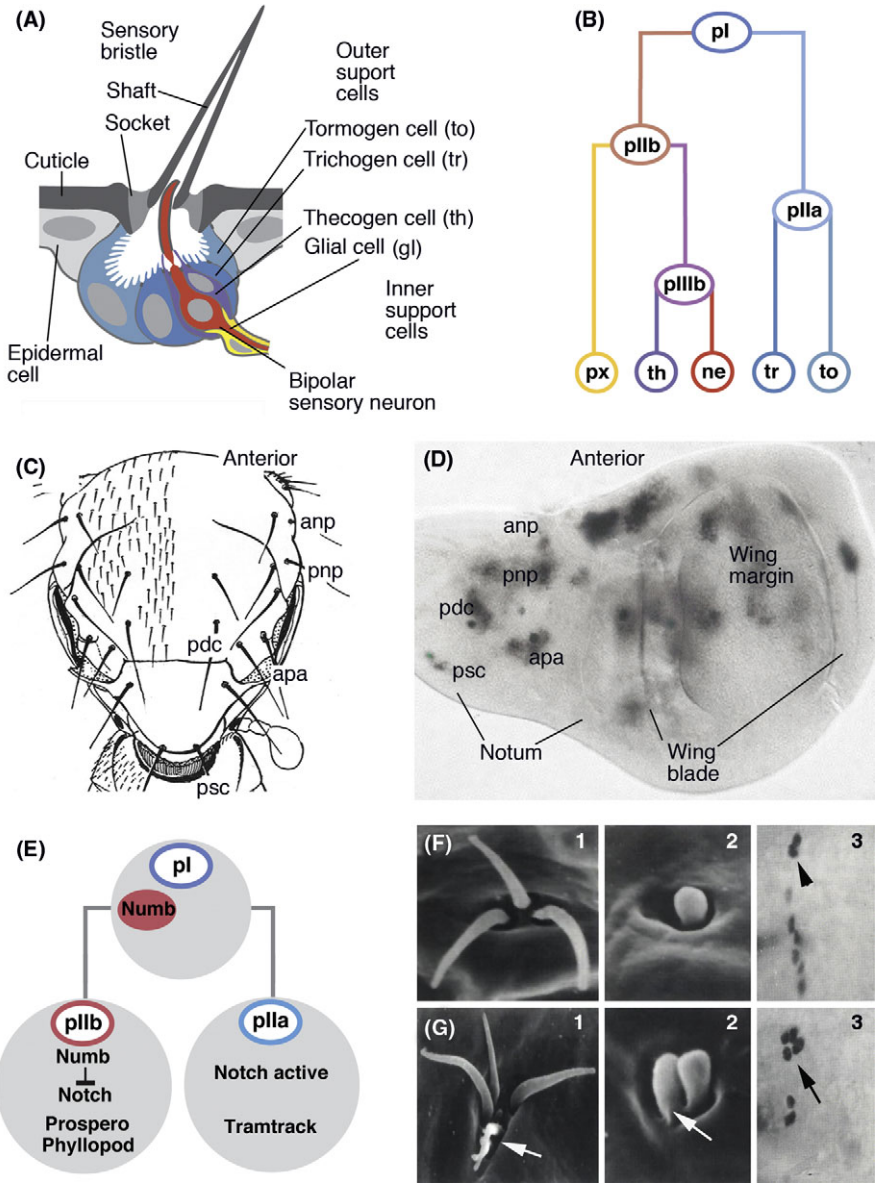
As the inner ear takes shape, the cells that form the otocyst wall diversify. In the same way that we learned about for the neural tube, different cell types delaminate from the epithelium and form additional layers basally adjacent to the epithelium (Fig. 14.6D). The anterior part of the otocyst forms a neurogenic region that sequentially, in two phases, produces the sensory neurons and hair cells of the inner hair. During the first phase of neurogenesis, expression of the proneural genes Neurogenin 1 and NeuroD promotes the formation of sensory neurons which innervate the cochlea and semicircular canals (Fig. 14.6E, H). During the second phase, cells of the neurogenic part of the otocyst switch on another proneural gene, Ath1 (homolog of the *Drosophila* atonal gene; see below), which prompts cells to adopt the fate of sensory hair cells (Fig. 14.6F, H). Cells of the neurogenic region that become neither sensory neurons nor hair cells will form support cells that anchor the hair cells in the epithelium lining the inner ear.

During the first phase of neurogenesis, Notch pathway activity restricts the number of cells that become sensory neurons. Notch activity is also required during the second phase of neurogenesis to distinguish between hair cells and support cells. Thus, precursors of hair cells appear at regular intervals, separated by support cell precursors (Fig. 14.6I–L). A Notch dependent lateral inhibition mechanism operates to determine the spacing of hair cells. Thus, at an early stage, the Notch receptor and the Delta ligand are expressed in all cells of the otic epithelium. Subsequently, Delta becomes restricted to the developing hair cells, which thereby inhibit their neighbors from choosing the same fate; these neighbors become the support cells. Genetic studies, using mouse and zebrafish, show that loss of Delta in the developing inner ear results in an increased number of hair cells (Takebayashi et al., 2007) (Fig. 14.6K, L).

In *Drosophila* and many other invertebrates, sensory organs are comprised of small clusters of sensory neurons and support cells, which together are called sensilla (Fig. 14.7). Sensilla are scattered over the entire body surface of the larval and adult fly. The cells that compose each sensillum are clonal descendants of a single Sensory Organ Progenitor cell (SOP) (Lai & Orgogozo, 2004). The fate of sensory neurons and their support cells is determined by factors expressed in the SOP and delivered through a series of asymmetric divisions to the individual cells. In most cases, all cells of a sensillum are born in three consecutive divisions. The SOP, also called the primary precusor cell or pI, divides into two pII daughters called pIIa and pIIb (Fig. 14.7B). PIIa divides one more time and gives rise to the outer support cells, which form processes in the shape of hairs or bristles. PIIb gives rise to the sensory neuron and inner support cells in two consecutive divisions (Fig. 14.7B). The first step in sensillum development, the determination of SOPs, is controlled in a manner similar to that regulating neuroblast determination. Thus, the expression of proneural genes defines proneural clusters in the ectoderm that are competent to form SOPs (Fig. 14.7C, D). At the same time, proneural genes activate the Notch pathway, mediated by the neurogenic genes, which single out one SOP and prompt all other cells of the proneural cluster to abort neural development. The identities of individual SOPs are specified by different proneural genes. For example, *Poxn* and *Amos* are expressed and required in precursors of chemosensory receptors. *Atonal* (*ato*) is expressed and required in the proneural clusters that give rise to stretch receptors (chordotonal organs). Loss of *ato* function results in the absence of these cell types; likewise ectopic expression of *ato* in other SOPs induces the production of chordotonal organs.

Once determined, SOPs divide asymmetrically, using the same mechanism discussed above for neural progenitors of the CNS. A crucial factor in generating the asymmetry of SOP divisions is the basal complex member, Numb (Fig. 14.7E). After the the first SOP mitosis, the cell receiving Numb becomes pIIb. Numb disrupts the Notch signaling pathway in pIIb, allowing for the expression of genes that are required for sensory neurons, including Prospero and phyllopod. If Numb is deleted, Notch can act in both pII cells, which changes the fate of pIIb to that of pIIa,

FIGURE 14.7 Sensory organ development in *Drosophila*. (A) Structure of mature sensilla. Schematic section of external mechanoreceptor (microchaete). Bipolar sensory neuron (red) is surrounded by two inner support cells, the thecogen cell (purple) and glial cell (yellow). Two outer support cells, the tormogen cell (light blue) and trichogen cell (dark blue), form the socket and shaft of the bristle-shaped cuticular component of the sensillum. (B) Tree diagram of canonical cell lineage of mechanoreceptive sensillum (microchaete). Progenitor of sensillum (SOP or pI; top) divides into two daughter cells, pIIa and pIIb. Division of pIIa gives rise to outer support cells (trichogen and tormogen cell). PIIb divides into progenitor pIIIb that then produces sensory neuron and thecogen cell; the other daughter of pIIb develops as glial cell (as shown here for the microchaete) or expresses a number of other different fates (not shown). (C and D) Sensillum patterning. (C) Schematic dorsal view of *Drosophila* thorax (notum) indicating invariant pattern of mechanoreceptive bristles. Some of the large bristles (macrochaetes) are identified individually. Anp, anterior notopleural; apa, anterior postalar; pdc, posterior dorsocentral; pnp, posterior notopleural; psc, posterior scutellar. (D) Expression of the proneural gene *scute* (grey) in wing imaginal disc. In the proximal region of the disc which gives rise to the notum, *scute* is expressed in proneural clusters which can be assigned to individual macrochaetes. (E–J) Distribution of fate determinants and Notch signaling in the sensillum lineage. (E) During the pI mitosis, Numb, an inhibitor of Notch signaling, is distributed to pIIb. This activates determinants of neurons like Prospero and Phyllopod. Loss of Numb in pI causes transformation of pIIb into pIIa, with a concomitant doubling of outer support cells and absence of neuron/inner support cells. (F, G) Examples of this transformation are the mechanoreceptive Keilin's organ that, in wild type (F1), has three sets of shaft/socket, and, in numb mutant (G1), has four to five sets of these outer support cells. Likewise, duplication of shaft/socket occurs in a basiconical sensillum (F2: wild-type; G2: numb mutant). F3 shows molecular marker Cut expressed in two outer support cells of a mechanoreceptive sensory organ of the embryo (arrowhead). In the numb mutant (G3), these cells are doubled in number (arrow).

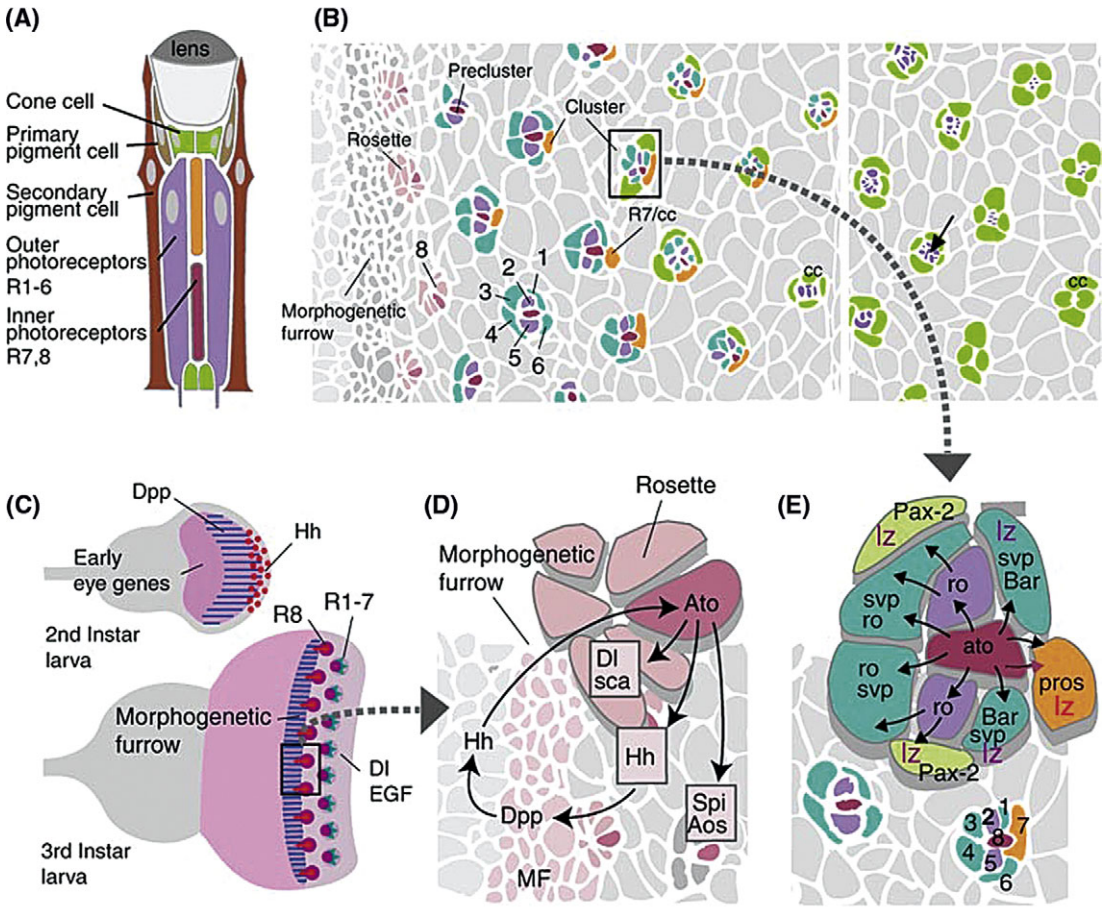


and causes a duplication of outer support cells (the descendants of pIIa) and loss of neuron/inner support cells (the descendants of pIIb; Fig. 14.7G). It is clear that recurring themes, such as spatially restricted homeobox containing transcription factors, proneural bHLH genes, asymmetric cell divisions, and Notch signaling play a role in sensory neuron determination in vertebrates and invertebrates just as they do in the central nervous system of these animals.

THE RETINA

The retina is a sensory organ with different origins. The *Drosophila* retina arises from an imaginal disc, whereas the vertebrate retina arises from an outpocketing of the anterior neural plate. Studies of how specific neurons arise and diversify in the retinas of these distinct animals reveal surprising similarities despite their very different embryological origins, betraying a shared molecular evolution of visual organs that predates the separation of vertebrates and invertebrates (Lamb, Collin, & Pugh, 2007).

The insect compound eye is built of a large number of identical facets, called ommatidia. *Drosophila* possesses approximately 800 ommatidia in each eye. Each ommatidium contains eight identifiably unique photoreceptor neurons and several accessory cells. Six of the photoreceptors (R1–R6) form an outer trapezoidal array, whereas two (R7 and R8) are located in the center (Fig. 14.8A). Among the accessory cells are cone cells, which form the lens of each ommatidium, and pigment cells, which surround the photoreceptors and optically shield the ommatidia from one another.



Copyright © 2002, Elsevier Science (USA). All rights reserved

FIGURE 14.8 Determination of ommatidial cell fate in the *Drosophila* compound eye. (A) Schematic longitudinal section of ommatidium depicting cell types in different colors. (B) Surface view of part of the eye imaginal disc at a stage when photoreceptor clusters become assembled in a sequential fashion. Gray profiles indicate the apical surface of undifferentiated cells; colored profiles demarcate different types of photoreceptors and cone cells. At the left margin, all cells are undifferentiated. This is followed by a phase in which eye disc cells become more closely packed and constricted at their apical pole, thereby forming the morphogenetic furrow. As they leave the furrow, cells form more or less regularly spaced rosettes (light red). Within each rosette, a single cell becomes singled out as the photoreceptor R8 (dark red). At the time when it becomes distinguishable morphologically, four other photoreceptors (R2, R5: lilac; R3, R4: blue) have joined R8. Together, these five cells form the so-called preclusters. Three more photoreceptors (R1, R6: blue; R7: orange) join the preclusters, leading up to complete photoreceptor clusters. This step is followed by the appearance of four cone cells (green) that surround the photoreceptors in a circular fashion. On the right side of the panel, photoreceptors can be seen to segregate from the surface. Their apical membranes (all shown in purple) become increasingly smaller (arrow) and finally disappear altogether. (C) Initiation of compound eye development. During midlarval stages (Top), the signal Hedgehog (Hh; red) is expressed at the posterior tip of the eye disc. Hh activates the expression of another signaling protein, Dpp (blue), which turns on several "early eye genes," which commit the undifferentiated, proliferating cells that comprise the eye disc to an eye fate. During late larval stages, Hh and other factors initiate differentiation of the first ommatidial cells, the R8 photoreceptors, by turning on the proneural gene atonal (*ato*). This sets in motion other signaling pathways (EGFR signaling; Notch/Delta signaling) required for the proper spacing of ommatidia and the determination of other ommatidial cell fates. (D) Determination and spacing of R8 are controlled by the proneural gene *atonal (ato)*, as well as the inhibitory signals Delta (Dl) and Scabrous (Sca). Initially switched on in the entire morphogenetic furrow, *ato* expression then becomes restricted to a mosaic of regularly spaced cells, which subsequently differentiate as R8. Continued secretion of Hh, which signals across the MF toward the more anterior cells of the eye disc, drives the morphogenetic furrow across the eye disc. (E) Determination of photoreceptors R1–R7 and cone cells. Secretion of the TGF β homologue Spitz (Spi) from R8 stimulates cells surrounding R8 via the *Drosophila* EGF receptor homologue (DER), the Ras signaling pathway (symbolized by black arrows). In conjunction with other signaling events, including Notch/Delta, photoreceptors are induced to express specific transcription factors, which are required for their respective fates. The *lozenge* (*lz*) gene is expressed in R1, R6, R7, and cone cells; *seven-up* (*svp*) is expressed in R1, 6, 3, and 4; *rough* (*ro*) in R2, 3, 4, and 5; *BarI* in R1, and 6; *prospero* (*pros*) in R7, and *Pax-2* in cone cells.

Cells of the ommatidia are formed from clonally unrelated, uncommitted precursor cells that are generated from the proliferation of cells in the eye imaginal disc (Mollereau & Domingos, 2005). The ommatidial cells do not appear all at once, but follow a reproducible temporal sequence (Fig. 14.8B). During late larval life, a wave of differentiation passes over the eye disc in a posterior-to-anterior direction. In front of the wave, cells are still dividing

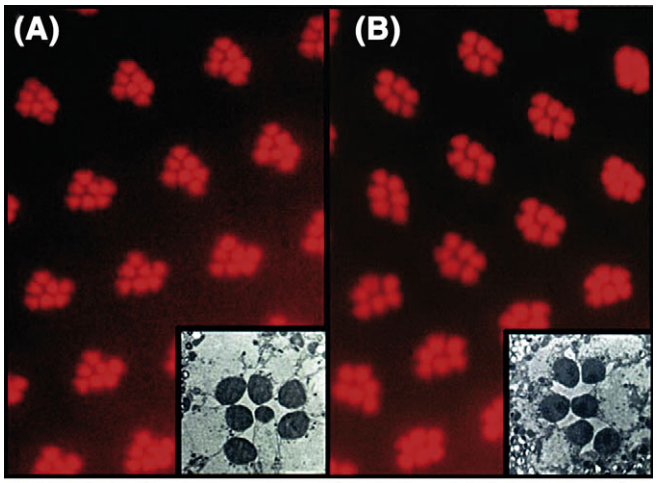
and uncommitted, whereas behind the wave front, which is visible as a depression in the disc called the "morphogenetic furrow" (MF), cells are mostly postmitotic and are beginning to differentiate. As they leave the MF, cells of the eye disc become arranged in a pattern of "rosettes" that foreshadow the regular ommatidial pattern. Soon afterward, one cell is singled out in each rosette to become R8. Being the first cells to differentiate, R8 cells can be considered as "crystallization centers" around which the other ommatidial cells aggregate in a stereotyped pattern.

R8 cells express the proneural gene *atonal (ato)*. Atonal is turned on by the signaling protein Hedgehog (Hh, a homologue of vertebrate Sonic hedgehog, Shh). Atonal thus first turns on in a continuous band of cells, similar to a preneural cluster, within the morphogenetic furrow, which then initiate lateral inhibition through Delta Notch signaling (Fig. 14.8C, D). By this mechanism, *ato* expression becomes restricted to a set of spaced cells, which subsequently differentiate as R8 photoreceptors. Atonal expression causes Hh to be produced and released. Thus, these R8 precursors continue emitting Hh, which signals across the morphogenetic furrow toward the more anterior cells of the eye disc to induce the next set of R8 cells (Ma, Zhou, Beachy, & Moses, 1993). This Hh-mediated feedback mechanism drives the wave of differentiation—that is, the morphogenetic furrow, across the eye disc, leaving in its wake a beautifully spaced mosaic of R8s, each crystallizing an ommatidium. Deleting *ato* results in the total failure of retinal development, which indicates that R8 plays a pivotal inductive role, triggering the determination of other cells that join the ommatidial clusters. The first cells to join the R8 cell shortly after its determination are R2, R3, R4, and R5 (Fig. 14.8A). Together with R8 they make up an ommatidial "precluster." The next cells to join the cluster are R1 and R6, followed by R7 and four cone cells; the last cells to join the cluster are the different types of pigment cells.

Shortly following its own determination, R8 puts out signals that activate two different signaling pathways: Notch and Ras. The Notch pathway, as we know, is activated by Delta. The Ras pathway represents a highly conserved biochemical cascade of cytoplasmic kinases (Ras, Raf, MPK), which in this case are activated by the epidermal growth factor receptor (EGFR). R8 emits a signal, Spitz (Spi), that activates this receptor and another signal, Argos (Aos), that acts as an inhibitor. Activation of these signaling cascades spreads concentrically from R8 to the precluster cells (R2, R3, R4, and R5) and then the remainder of the ommatidial cells (Fig. 14.8D). The precise, temporally regulated activation of EGFR and Notch signaling pathways assigns distinct phenotypes to the cells that join the ommatidial clusters.

The way in which differential signaling pathway activation controls ommatidial cell fates has been studied most closely for two cell types, R7 and the cone cells, which appear at a relatively late stage when all other photoreceptors are already in place. The *Pax-2* gene specifies cone cells (Fig. 14.8E). To express *Pax-2*, the cone-cell-to-be depends on three inputs (Flores et al., 2000): (1) the previous expression of the transcription factor lozenge (*lz*); (2) activation of the Notch pathway; and (3) activation of the EGFR signaling pathway. If any one of these prerequisites is not met, the cells will fail to express *Pax-2* and will not differentiate as cone cells. Conversely, if a set of cells that normally does not receive all three inputs is exposed to them experimentally, these cells will express *Pax-2* and become cone cells. R7 cells normally express *lz*, but do not activate the Notch pathway, and experimental misexpression of activated Notch in the R7 precursor will turn this cell into a cone cell.

The determination of the R7 cell deserves special mention in view of the pivotal role it has played in opening up the molecular-genetic study of signaling pathways (Zipursky & Rubin, 1994). The important advantage a genetic model system like *Drosophila* has to offer is that developmentally relevant molecules, such as signaling proteins, receptors, and signal transducers, can be identified by mutant screens. One of the first of such screens in the field of cell determination took advantage of the fact that only one of the photoreceptor cells of each ommatidium, R7, is sensitive to ultraviolet (UV) light. Thus, mutagenizing flies and screening for offspring that are blind to UV light yielded a fly line that lacked the R7 cell in every ommatidium and was therefore aptly called *sevenless* (Harris, Stark, & Walker, 1976) (*sev*) (Fig. 14.9). The close study of this gene showed that it encodes a receptor with



Copyright © 2002, Elsevier Science (USA). All rights reserved.

FIGURE 14.9 Photoreceptors in the eye of normal and *sevenless* mutants. (A) If a light is shone from the back of a fly's head and focused in the facets of the eye, individual photoreceptors can be seen because of their ability to pipe light. The wild-type animal has the normal pattern of seven photoreceptors visible in each facet. The small one in the center is photoreceptor R7. (B) The same technique used in a *sevenless* mutant shows only the six large photoreceptors R1–6 in each facet. R7 is missing. Insets show electron micrographs through single facets.

tyrosine kinase activity and that this receptor is expressed (among other cells of the eye disc) in the cell that will become R7. Lack of the receptor causes the cell that would normally become R7 to develop as a cone cell instead. In several follow-up screens, many signal transducing molecules and transcription factors activated by the Sev receptor were identified. Among them were the *Drosophila* homologues of Ras, Raf, and MAPK. One of the most rewarding findings was the identification of the signal that binds to Sev. Given the widespread expression of Sev, it was clear that a signal emanating from a point source must exist, in order to ensure that Sev would become active only in one cell in each ommatidium. The logical candidate to emit such a signal would be R8, the only cell in broad contact with the presumptive R7. A genetic screen for R7-less flies yielded identification of a membrane bound signaling molecule, called Bride of sevenless (Boss), which is expressed specifically in R8 cells and serves as the ligand for Sev (Reinke & Zipursky, 1988). These experiments provided an impressive demonstration of the awesome power of genetic screens combined with precise knowledge of the developing system in identifying new genes and proteins controlling development.

The vertebrate retina, like the *Drosophila* retina, develops from a population of pluripotent neuroepithelial progenitors, which produce a diversity of neurons and glia (Fig. 14.10A). As in *Drosophila*, vertebrate neurogenesis and determination in the retina follows a temporal order (Livesey & Cepko, 2001). Thus, at any given time during retinogenesis, only a few fates are available to differentiating cells. Such temporal fate restrictions are likely to be caused by transiently expressed transcription factors that switch on over particular periods of neurogenesis, such as Ikaros, a vertebrate homologue of Hb, which plays a similar role in *Drosophila* histogenesis. Cells born early in the vertebrate retina generally adopt fates as retinal ganglion cells (Fig. 14.11A–C), followed sequentially by horizontals and amacrines, bipolars, photoreceptors, and Müller glia. This is also similar to *Drosophila* in which first R8s and then sequentially other cell types differentiate. This progressive shift in cell type genesis in vertebrates is supported by dissociation experiments in which cells are put into culture at low density at various stages of development. If progenitors are isolated at the time when retinal ganglion cells (RGCs) are normally born, they tend to turn into ganglion cells in culture. Later progenitors in the same culture medium become mainly photoreceptors (Watanabe & Raff, 1990).

There is an intrinsic element to the succession of cellular fates in the vertebrate retina (Cayouette, Poggi, & Harris, 2006), which is revealed when rat retinal progenitors are isolated and allowed to divide on their own in culture. These isolated progenitors give rise to clones of cells of the same general birth order, size, and composition as clones *in vivo*. If one follows these clones as they develop, there is another surprising similarity to *Drosophila* retinal determination, which is that cells seem to choose their cell type in a stochastic manner—that is, lineage does not predict fate.

So how do vertebrate retinal cells become determined so that there are the right number of all the different cell types? As in *Drosophila*, vertebrate homologue of *ato*, *ath5*, is turned on early, when the first retinal cells are being born. Delta-Notch signaling then ensures that not too many cells become RGCs. If *ath5* expressing progenitors are followed by time-lapse analysis in the zebra fish retina (Poggi, Vitorino, Masai, & Harris, 2005), they appear to divide to generate one RGC that expresses ath5 at high levels and one other cell in which ath5 levels are reduced (Fig. 14.10B). Several other transcription factors are expressed at about this time in various populations of cells as they begin to differentiate. One such factor, Ptf1a, is turned on in the subset of newly postmitotic cells that will become inhibitory neurons in the retina (i.e., the amacrine and horizontal cells).

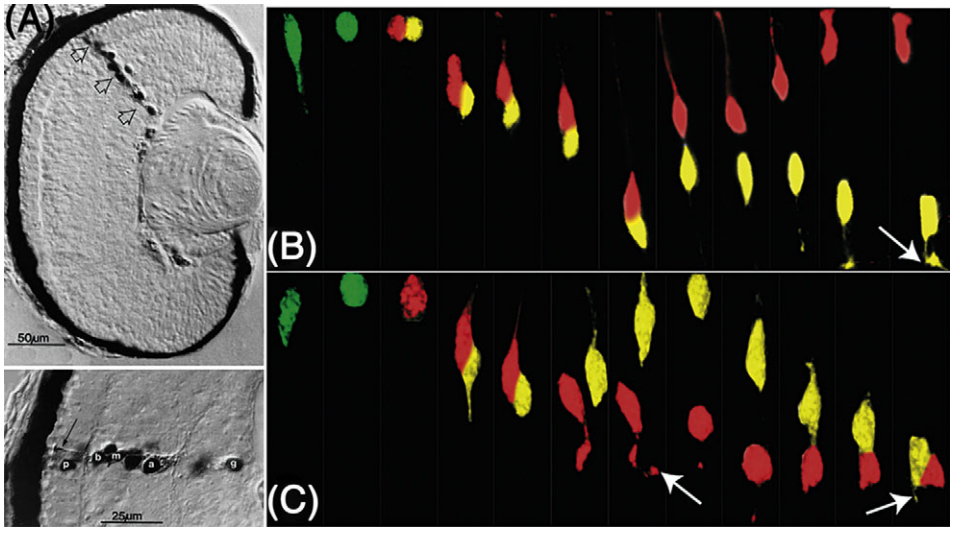


FIGURE 14.10 Clones of cells in the *Xenopus* and zebra fish retina. (A top) Daughters of a single retina progenitor in *Xenopus* injected with horseradish peroxidase are seen to form a column that spans the retinal layers and contributes many distinct cell types, (A bottom) p, photoreceptor; b, bipolar cell; m, Muller cell; a, amacrine cell; and g, ganglion cell. (B) Time-lapse images of an RGC progenitor in zebra fish dividing at the apical surface of the retina to produce two cells one of which (yellow) becomes an RGC. (C) A similar RGC progenitor in a mutant retina where there are no host RGCs because it is mutant for the *ath5* gene. This cell divided to produce two daughters, both of which become RGCs (arrows point to emerging axons).

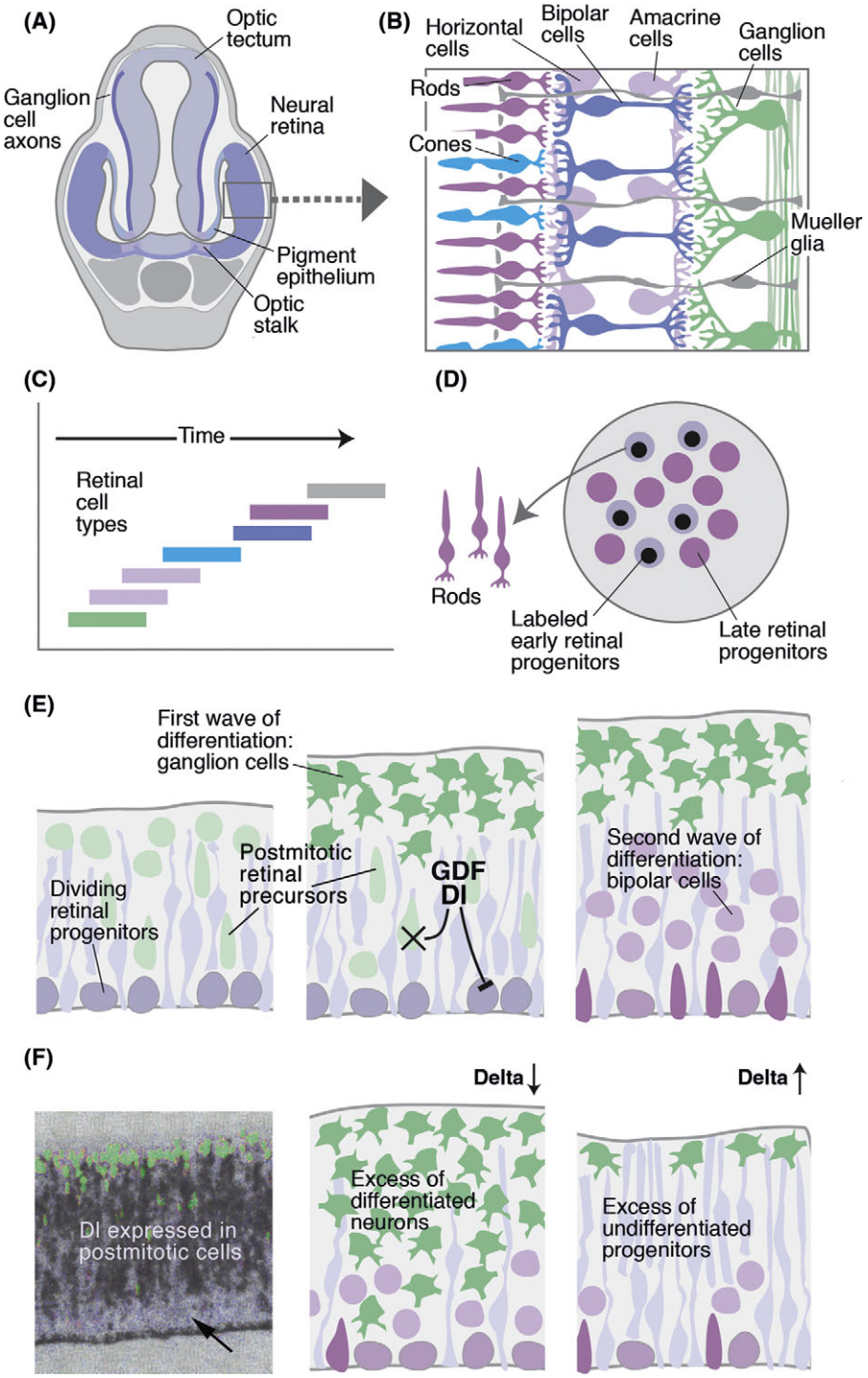


FIGURE 14.11 Vertebrate retina development. (A) The neural retina bulges out of the ventral neural tube at the level of the diencephalon. It is joined to the brain by the optic stalk along which the axons of retinal ganglion cells will course on their way to the tectum. (B) The seven major cell types in the retina coded by color. Their laminar arrangement by cell type is evident. (C) Birth dating studies in the retina show that different cell types are born in different but overlapping periods of development. (D) When early generated retinal precursor cells (labeled by a pulse of BrdU) are mixed with older cells in culture, they show an increased probability of turning into late cells, such as rods. (E) Development of the retina. Schematic section of retinal neuroepithelium at three different stages (early: left; late: right). Dividing retinal progenitors (purple) produce different cell types depending on the developmental stage. Differentiation inhibitors (Dl) released from postmitotic neural precursors inhibit retinal progenitors from producing more ganglion cells. Postmitotic cells also secrete differentiation factors such as GDF which limit the time window during which postmitotic cells can become retinal ganglion cells. (F) Expression and function of Dl in the developing retina. Left panel: expression of Dl in postmitotic neural precursor; dividing progenitors of the ventricular layer (arrow) or differentiated neurons of the basal mantle layer (green) do not express Dl. Middle: Loss of Dl results in increased number of early-born retinal cells (retinal ganglion cells), at the expense of undifferentiated retinal progenitors. Right: Overexpression of Dl causes the opposite phenotype.

Ath5 is necessary for the determination of RGCs, while Ptf1a is essential for the generation of inhibitory neurons. Ptf1a also controls inhibitory neuron differentiation in the spinal cord (see the following section). In the absence of Ath5 cells that would have become RGCs and in the absence of Ptf1a, cells that would have become amacrine or horizontal cells choose excitatory fates instead. All horizontal cells, it turns out, express both Ath5 and Ptf1a. This begins to look like a combinatorial code. Indeed, other studies have shown that particular combinations of retinal transcription factors, sometimes two or three together, drive cells to become specific cell types, while single factors are much less effective (Ohsawa & Kageyama, 2008).

How these different intrinsic determinants are turned on in these combinations is unknown, but it is clear that fate determination in the retina is not just an intrinsic affair.

When retinal cells are removed at the stage when RGCs are being born, mixed into aggregates, and cultured *in vitro*, they differentiate into RGCs as one would expect. However, if the same cells are mixed with an excess of retinal cells several days older (i.e., when photoreceptors are generated), they generally become photoreceptors (Fig. 14.11D). This work shows that individual cells have the capacity to differentiate into different cell types, and the fate they choose depends on the environment in which they are born. This makes good sense, for as retinal development proceeds, the environment changes simply as a result of various retinal cell types being generated, much as in the *Drosophila* retina. Indeed, there is strong evidence for negative feedback in influencing the fate of RGCs. Young embryonic chick progenitors are inhibited in their ability to produce RGCs when cultured adjacent to older retinas in which there are lots of RGCs, and depletion of the RGCs from these older retinal cell populations abolishes this inhibition (Waid & McLoon, 1998). When normal retinal progenitor cells that express a fluorescent protein under the control of the *ath5* promoter are transplanted to normal embryos, the cells that turn on ath5 usually generate only a single RGC daughter (see above), but these same cells when transplanted into mutant retinas that lack RGCs, they often generate two RGC daughters, providing evidence that such lineages are sensitive to feedback signals (Fig. 14.10C) (Poggi et al., 2005). It seems that Ath5 itself is negatively regulated by such feedback signals. Similarly normal progenitors transplanted into Ptf1a mutant retinas are more likely to turn on Ptf1a to try to compensate for the missing cell types, suggesting again that negative feedback normally keeps the level of Ptf1a expression appropriate to generate the correct proportion of inhibitory neurons in the retina (Jusuf et al., 2011).

What could these feedback signals be? First, as mentioned above, the Notch signaling pathway contributes to retinal diversity in the vertebrate retina (Fig. 14.11E, F). When Notch signaling is compromised in early retinal progenitors, too many cells differentiate as early cell types, whereas later interference with Notch signaling leads to an increase in later-born cell types. The Notch signaling pathway, by allowing only a certain number of cells to express particular proneural genes and thus differentiate at any one time, is important in creating cellular diversity in the vertebrate retina (Dorsky, Chang, Rapaport, & Harris, 1997). RGCs also express Shh, and the Shh knockout mouse contains increased number of RGCs. These results suggest that Shh provides a feedback inhibition signal in the developing mouse retina, echoing what happens in the *Drosophila* retina where R8 cells express Hh, which regulates the expression of *ato*. The growth differentiation factor 11 (GDF11) is another factor secreted by RGCs, and it acts by limiting the temporal window during which progenitors are competent to express *ath5* and thus produce RGCs.

THE SPINAL CORD

The vertebrate spinal cord plays a central role in conveying motor commands to the periphery via somatic and autonomic motor neurons. It also harbors neurons that receive and transmit cutaneous, proprioceptive, and visceral sensory information from specialized receptors in the body and limbs. This information is used to activate protective reflexes such as the nociceptive withdrawal reflex, and to shape motor activity needed for locomotion. The component neurons of these spinal sensorimotor circuits are generated by the patterning events that specify neuronal identity in the embryonic spinal cord. Antero-posterior positioning contributes to neural patterning in the cord as evidenced by differences in the number and types of specialized cells that are found at different axial levels. This is most apparent for branchiomotor, phrenic, preganglionic, and lateral motor column motor neuron cell types that are patterned by the Hox genes (Dasen, Liu, & Jessell, 2003; Dasen, De Camilli, Wang, Tucker, & Jessell, 2008). The Hox genes are also likely to regulate the diversity and abundance of specialized interneuron cell types that are similarly distributed in different numbers along the neuroaxis.

The dorso-ventral position in which a neuronal progenitor is located within the neural tube is one of the major determinants of neuronal cell fate in the spinal cord. Signaling centers at the ventral and dorsal midline establish opposing gradients of Shh and BMP signaling, respectively. As a result of this signaling the ventricular zone is subdivided into many discrete progenitor domains that in turn produce motor neurons and all the classes of spinal interneuron. The ventral-most region of the spinal cord, the "floor plate" is induced by a Shh signal emitted from the underlying notochord (Chapter 13). Once induced, the floor plate serves as a secondary source of Shh, which percolates dorsally through the cord to provide graded Shh signaling to ventral neural tube progenitors (Fig. 14.12A). Differential Shh signaling partitions the ventral basal plate into five discrete progenitor domains that in turn generate six neuronal cell types (Jessell, 2000).

How do cells at different dorso-ventral levels interpret their exposure to different levels of Shh in order to acquire different fates? Threshold levels of Shh activate homeobox transcription factors such as Nkx2.2 and Nkx6.1 (Class II) ventrally, while at the same time turning off other factors such as Pax6, Dbx2, and Pax3/7 (Class I) dorsally (Goulding, Lumsden, & Gruss, 1993; Jessell, 2000). Thus, neural tube progenitors in different dorso-ventral regions express specific homeodomain proteins, which either are turned on or off at particular Shh thresholds (Fig. 14.12B). The boundaries between these domains are sharpened through reciprocal repression of the two classes of transcription factor. The ventral border of the class I Pax-6 gene initially overlaps the dorsal border of the class II

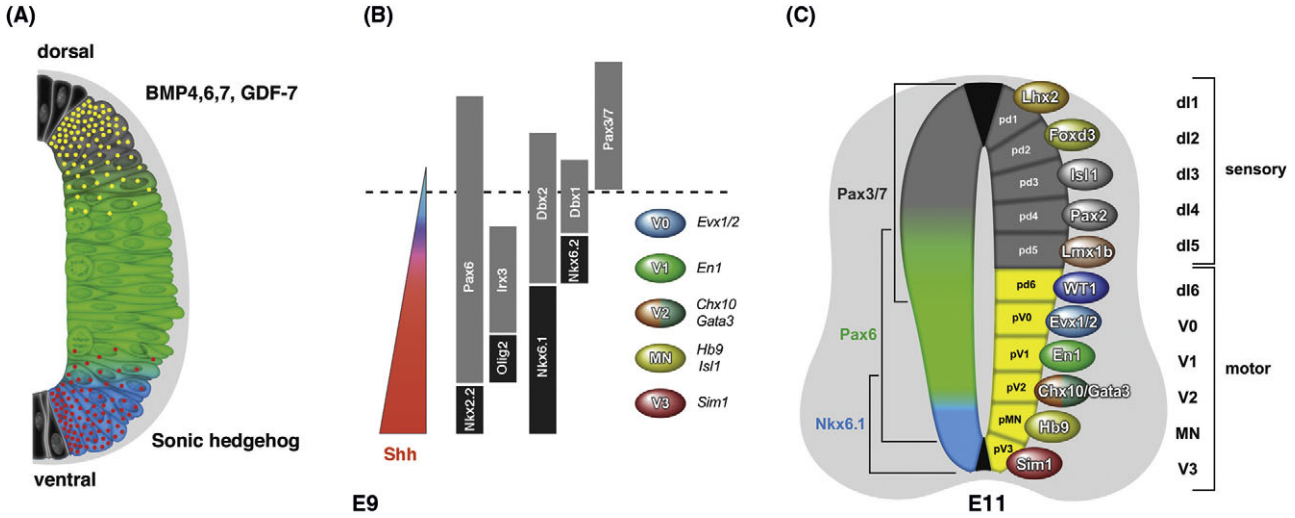


FIGURE 14.12 Specification of neurons in the vertebrate spinal cord. (A) Schematic cross section of the mouse neural tube. Shh protein (red) is released ventrally from the notochord and floor plate (black cells). It forms a gradient with high concentrations ventrally and low concentrations dorsally. BMPs and other TGF-B-related molecules are released from the dorsal epidermis and dorsal neural tube to form an opposing gradient. These opposing gradients lead to the expression of Nkx6.1 (blue) ventrally, Pax6 (green) medially, and Pax3/Pax7 (dorsally). These expression patterns are refined further (see C) and they contribute to the subdivision of the ventricular zone (inner band of cells) into eleven distinct progenitor domains (pd1–pd6 and pV0 –pV3). (B) A Shh gradient (red) in the ventral half of the neural tube establishes the expression domains of the class I and class II homeodomain genes (see text for details). This in turn leads to the development of five ventral progenitor domains that give rise to V0, V1, V2, MN, and V3 neurons. Cross-repressive interactions between class I (gray) and class II (black) homeodomain genes establish the boundaries between these progenitor domains. (C) Representative cross section through the mouse spinal cord at E11.5 showing the primary classes of neuron that are generated in the embryonic spinal cord. Neurons that arise ventrally (dI6 –V3) are thought to become components of the spinal motor circuitry, while more dorsal cell types (dI1–dI5) contribute to the reception and transduction of sensory information from the body.

Nkx2.2 gene, but then reciprocal repression sets in, so that only one of these factors is expressed in any particular progenitor cell. By this process, specific domains uniquely express particular combinations of Class I and Class II homeodomain transcription factors.

Each domain that is generated by cross-repression produces a single generic class of neuron. V3 interneurons that arise close to the floor plate are generated from progenitors that express Nkx2.2, but not Pax6 or Olig2. Motor neuron progenitors express *Nkx6.1* but not *Irx3* and *Nkx2.2*. *Nkx6.1* turns on Olig2, a bHLH transcription factor that is required for motor neuron differentiation, and Olig2 in turn activates the expression of motor neuron specific transcription factors, such as HB9, Lim3, and Isl1/2, that are each involved in motor neuron differentiation (Fig. 14.12B, C).

A variety of interneuron cell types are generated from the other progenitor domains, with each domain producing a single primary generic class of interneuron (Fig. 14.12C). The V2 progenitor domain is unique in that it generates two interneuron classes: V2a and V2b INs. Notch signaling controls the specification of these two cell neuronal types, which results in the production of equivalent numbers of V2a and V2b neurons (Peng et al., 2007). Dll4 is selectively expressed in a subset of V2 progenitors and these cells activate the Notch pathway in sibling V2 neurons, which in turn induces the expression of Scl and the V2a differentiation program (Del Barrio et al., 2007). The V1 interneuron class that arises just dorsal to the V2 progenitor domain exemplifies many of the features of these early interneuron classes. While these cells share a number of similar attributes, rostral directed ipsilateral axons and inhibitory neurotransmitter phenotype, they are comprised of multiple specialized cell types, possibly five or more. This specialization of neuronal phenotype appears to be regulated by a strict temporal program that overlays the DV spatial program (Stam et al., 2012).

In the dorsal embryonic cord, the alar plate can be coarsely subdivided into two major progenitor regions: a more dorsal region of the neural tube comprising three progenitor domains that generate Class A dI1–dI3 neurons and a second intermediate region that generates Lbx1-expressing Class B neurons (Fig. 14.13). Class A neurons (dI1–dI3) are generated in a BMP and Wnt-dependent manner, whereas Class B (dI4–dI6) interneuron differentiation is independent of BMP/Wnt signaling. The downstream targets of the dorsal BMP/Wnt pathway include the bHLH transcription factors Olig3, Atoh1, Neurog1/2, and Ascl1, which function as cell type specific determinants in Class A progenitors. Olig3 is expressed in all Class A progenitors, where it is required for dI1–dI3

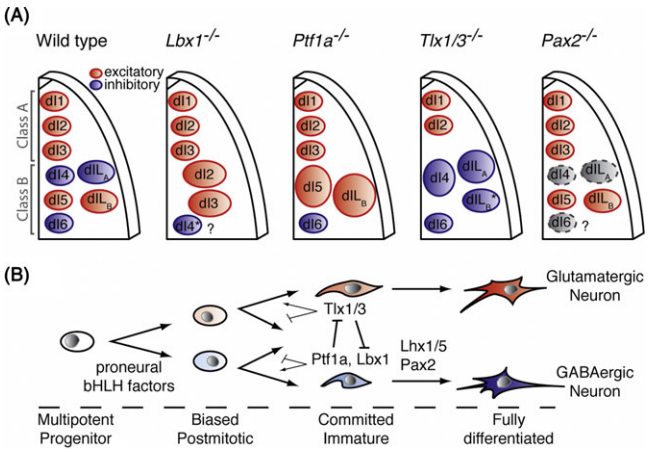


FIGURE 14.13 Genetic control of neuronal cell types and neurotransmitter phenotype in the dorsal spinal cord. (A) Six early (dI) and two late (dIL) populations are generated in the embryonic spinal cord. dI1–3, dI5, and dIL_B neurons are excitatory glutamatergic neurons (red), and dI4, dI6, and dIL_A differentiate into inhibitory cell types (blue). From left to right: In *Lbx1* mutant mice, Class B neurons fail to form, and cells are converted to Class A fates. Excess dI2–dI3 glutamatergic cells are generated. In *Ptf1a* mutant mice, dI4 and dIL_A interneurons transfate to glutamatergic dI5 and dIL_B neurons. This results in the complete loss of inhibitory neurons in the dorsal spinal cord and an excess of excitatory neurons. In *Tlx1/3* mutants, dI3, dI5, and dIL_B excitatory interneuron formation is impaired, and there is a concomitant upregulation of GABAergic markers. Mutant dIL_B cells (*) maintain Lmx1b expression, but they switch to an inhibitory phenotype. In *Pax2* mutants, inhibitory neuron markers are lost, but dI4/dIL_A neurons do not adopt the identity of excitatory dorsal neurons. (B) Model for the specification of excitatory or inhibitory transmitter phenotype in the dorsal spinal cord. Crucial genetic interactions occur in early postmitotic neurons between HD and bHLH transcription factors. Gain or loss of these factors produces a fate switch, with additional factors such as Pax2 operating in postmitotic neurons to consolidate the neurotransmitter program.

neuron specification. It does so by repressing the expression of Lbx1 (Muller et al., 2005), an essential postmitotic regulator of Class B neuronal identity. Within the Class A region, the proneural bHLH transcription factors Atoh1, Neurog1/2, and Ascl1 function at two levels. They act as proneural determinants for dI1, dI2, and dI3 neurons, respectively, and they regulate boundary formation between the dI1, dI2, and dI3 progenitor domains via a cross repression mechanism (Gowan et al., 2001), thereby subdividing the Class A region into three distinct progenitor domains.

A number of homeodomain transcription factors, including Pax3/Pax7, Msx1/3, and Gsx1/Gsx2, also have essential roles in patterning the alar plate. Pax3 and Pax7 are broadly expressed in alar plate progenitors where they appear to act as general determinants of dorsal identity (Fig. 14.12C). Gsx1 and Gsx2 are expressed in dI3–dI5 progenitors, where they control the specification of dI3, dI5, and dIL_B neurons (Mizuguchi et al., 2006) by upregulating the expression of Ascl1, which is an essential proneural determinant for all three cell types. Gsx2, which is required for the differentiation of dI3 interneurons, promotes Ascl1 expression in dI3 progenitors by blocking the expression of Neurog1/2. This suggests that homeodomain patterning factors like Gsx2 promote specific differentiation programs by functioning as permissive factors for proneural determinants such as Ascl1.

Class B Lbx1-expressing neurons come in two flavors: GABAergic inhibitory interneurons and excitatory glutamateregic interneurons. These two populations are marked by the mutually exclusive expression of Pax2, Lhx1/5, Gbx1 (inhibitory dI4 and dIL_A neurons), or Lmx1b, Tlx1/3 (dI5 and dIL_B excitatory neurons). This dichotomy suggests neurotransmitter choice and neuronal identity are closely linked. Indeed, the aquisition of a glutamatergic phenotype depends on the activity of the Tlx1 and Tlx3 selector HD transcription factors that determine dI5 and dIL_B identity (Cheng et al., 2004). Tlx1/3 in turn represses the Pax2-dependant inhibitory differentiation program. Conversely, the bHLH factor Ptf1a drives Class B neurons to differentiate as GABAergic INs by repressing Tlx1/3 expression (Glasgow, Henke, MacDonald, Wright, & Johnson, 2005; Mizuguchi et al., 2006). While the Tlx1/3 and Ptf1a transcription factors appear to be the primary determinants of Class B neurotransmitter phenotype, Pax2 and Lhx1/5 function downstream of Ptf1a to consolidate the GABAergic neurotransmitter program. Inactivation of the *Pax2* gene results in the loss of inhibitory neurotransmitter markers from dorsal dIL_A neurons. In this instance, Tlx1/3 and glutamatergic markers are not upregulated, which indicates that these cells do not convert to excitatory sensory neurons.

MOTOR NEURONS

Motor neurons in the spinal cord are organized in columns at different anterior-posterior levels. At limb levels, three major columns are evident: the medial motor column that innervates the axial muscles of the back, the hypaxial motor column that projects to the ventral body wall muscles, and the lateral motor column that innervates the appendicular musculature muscle (Dasen et al., 2003, 2008; Jessell, 2000). In the cervical region and lumbar spinal cord, the Lateral Motor Column (LMC) innervates forelimb and hindlimb muscles, respectively. In the midthoracic region, the Preganglionic Motor Column (PGC) innervates the sympathetic chain (Fig. 14.14A–C). A gradient of FGF and retinoic acid signaling selectively establishes differential Hox gene expression along the AP axis at different levels, boundaries between which are sharpened by reciprocal inhibition. This anterior-posterior logic of motor neuron identity is thus similar to that governing the positioning of the neurons in the dorsal ventral axis of

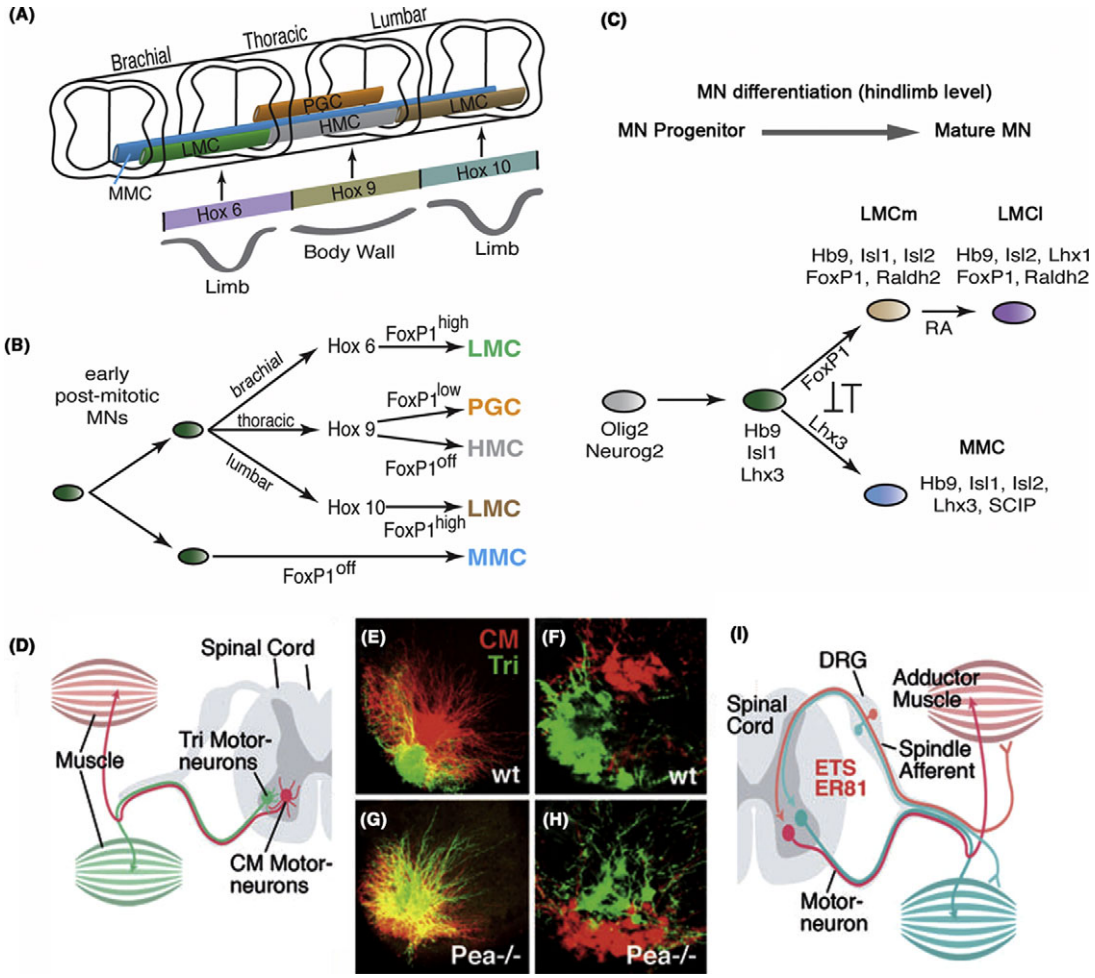


FIGURE 14.14 Establishment of motor neuron identity along the antero-posterior axis of the spinal cord. (A) Polarized expression of FGF and retinoic acid (RA) is responsible for the regionally specific expression of Hox genes along the antero-posterior axis of the neural tube. In addition, inhibitory interactions among the Hox genes themselves sharpen the boundaries within which a given Hox gene is expressed. Hox genes are differentially expressed along the AP axis of the spinal cord and function in conjunction with the forkhead domain gene FoxP1 to establish motor neuron columnar identity. Motor columns are patterned in a manner that corresponds to the arrangement of their targets at brachial, thoracic, and lumbar levels. The MMC innervates muscles of the body wall and therefore extends along the length of the spinal cord. The LMC innervates limb muscles and is present at brachial and lumbar levels. The HMC and PGC innervate muscles of the body wall and autonomic neurons of the sympathetic ganglia, respectively, at thoracic levels. (B) MNs acquire columnar identity based on differential expression of Hox6/9/10 and FoxP1 off/low/high along the AP axis. In *FoxP1* mutants, MNs assume MMC or HMC fates along the entire AP axis and other motor columns do not form. (C) Proposed model of the genetic specification of MN progenitors in lumbar cord. The actions of FoxP1, Lhx3, and retinoic acid (RA) establish the combinatorial expression profiles associated with each of the motor columns in lumbar spinal cord. (D) The motor column is subdivided into pools of motor neurons innervating specific muscles. These pools, including the Tri motor neuron pool and the CM motor neuron pool, occupy discrete locations in the ventral horn of the spinal cord and show distinct dendritic branching patterns. These properties, which depend on the expression of specific transcription factors such as the LIM gene PEA3, can be visualized by backfilling the motor neurons with fluorescent labels injected in the corresponding muscles. (E and F) Two different markers visualize the dendritic arborization (E) and cell body position (F) of the Tri motor neurons (green) and CM motor neurons (red). (G and H) Knockout of the PEA3 gene results in ventral shift of CM motor neuron location, and dorsal expansion of Tri motor neuron dendrites. (I) Reflex circuits between sensory afferents (from muscle spindles) and motor neurons innervating the corresponding muscle are formed because both spindle afferents and motor neurons targeted by them express a common transcription factor, such as ETS ER81 shown here. Among the downstream targets of ETS ER81 are homophilic adhesion molecules that allow the ingrowing sensory axon to "recognize" the proper motor neuron target. Abbreviations: MMC, median motor column; LMC, lateral motor column; HMC, hypaxial motor column; PGC, preganglionic autonomic motor column; RA, retinoic acid. *A–C adapted from Dasen et al. (2008); Rousso et al. (2008). E–H adapted from Vrieseling and Arber (2006).*

the spinal cord. FoxP1, which acts as a co-factor for the Hox transcription factors is required for the specification of LMC type motor neurons, and the PGC or Column of Terni as it is otherwise known (Dasen et al., 2008; Rousso, Gaber, Wellik, Morrisey, & Novitch, 2008).

The lateral motor columns are subdivided into pools of motor neurons that innervate different muscles. One major subdivision within the LMC is between motor neurons that innervate dorsal limb muscles and motor neurons that innervate ventral limb muscles. Transcription

factors of the LIM homeodomain (LIM-HD) family are involved in specification, as motor neurons expressing different combinations of LIM-HD factors such as Isl1, Isl2, Lhx1, and Lhx3 are expressed in different motor neuron classes that innervate different muscle groups (Fig. 14.14C). This is particularly well illustrated in the primary motor neurons (Rop, Mip, and Cap) of zebra fish, which can easily be distinguished by their relative rostral to caudal position in each spinal segment, as well as their axon trajectories to distinct muscle regions. These three neurons express different combinations of LIM transcription factors. The identity of these neurons, as judged by their projection to different muscle regions, is influenced by the position in which they develop. These projection patterns can be changed by transplantation in which the position of these neurons is changed prior to axon extension. Transplantations that lead to such switches in projection correlate with induced changes in LIM-HD code suggesting a causal link (Appel et al., 1995).

Motor neurons that innervate specific muscles are grouped into motor pools, which are distinguished by the expression of distinct members of the ETS family of transcription factors. For example, the ETS gene *ER81* is expressed in the motor neurons that innervate the limb adductor muscle in chicks, whereas the iliotrochanter motor neurons express the *PEA3* ETS gene (Fig. 14.14D–I). The initial expression of these ETS genes depends on peripherally derived signals such as GDNF and coincides with the arrival of motor neuron terminals in the vicinity of their specific targets. In PEA3 knockout mice, the motor neurons that are normally PEA3 positive fail to develop normally so muscles that are normally targets of these neurons are hypoinnervated and become severely atrophic. Not only do these motor neuron pools fail to branch normally within their target muscles, but the cell bodies and dendrites of these motor neurons also are mispositioned within the spinal cord, which is critical for the control of their reflexive and descending innervation (Fig. 14.14E–H). Recent work suggests that these ETS genes may regulate the expression of cell adhesion and axon guidance factors that help motor neurons recognize their target. Interestingly, the sensory afferents that innervate the stretch receptors in a particular muscle express the same ETS gene as the motor neurons that innervate that muscle, suggesting this may help the sensory neuron axons find the dendrites of these motor neurons to complete the connection that is the basis of the monosynaptic stretch reflex.

The motor neuron system in the ventral nerve cord of the *Drosophila* larva, which is responsible for crawling, displays a relatively simple repetitive organization (Landgraf & Thor, 2006). In each half segment, there are approximately 36 motor neurons arising from about 15 neuroblasts (Fig. 14.15A, B). Some project through the intersegmental nerve (ISN) to innervate the more interior muscles that span the length of each segment and cause the animal to shorten. Others project through the segmental nerve (SN) to innervate the more exterior set of muscles that are more radial in orientation and cause the animal to lengthen. This is somewhat comparable to the division in vertebrate limbs, where the sets of (usually antagonistic) muscles derived from the ventral and dorsal muscle masses are innervated by pools of motor neurons whose axons project along divergent paths. ISN and SN motor neurons are further subdivided into subclasses of motor neurons that innervate distinct muscle territories.

No master/selector gene has been identified that specifies generic motor neuron fate in *Drosophila*, although combinatorial codes of transcription factors specify different motor neuronal subtypes (Landgraf & Thor, 2006). The first subdivision in *Drosophila* is between ventrally projecting motor neurons, vMNs, and dorsally projecting motor neurons, dMNs. Interestingly, vMN identity is controlled by the combinatorial action of the *Nkx6* and *HB9* homeobox genes (Broihier & Skeath, 2002; Cheesman, Layden, Von Ohlen, Doe & Eisen, 2004), a similar combination to that used to specify generic motor neuron fate in vertebrates. The vMNs in *Drosophila*, like the motor neurons in vertebrates, are then specified into subclasses that project along specific nerves to innervate specific muscles, by the further expression of combinatorial codes of LIM-HD factors (Fig. 14.15C). dMNs depend upon other sets of transcription factors: even-skipped (eve) and grain (grn) (Garces & Thor, 2006; Landgraf, Roy, Prokop, VijayRaghavan, & Bate, 1999). In contrast to the LIM-HD factors, Nkx6 and HB9, the vertebrate orthologues of even-skipped (Evx1/2) and grain (Gata2/3) are not involved in vertebrate motor neuron specification. Thus, *Drosophila* vMNs appear to be evolutionary more similar to vertebrate motor neurons.

What aspects of motor neuron phenotype do combinatorial codes of transcription factors control? This issue has remained elusive, but more recent studies have pointed to two important motoneuron properties: axon pathfinding molecules and electrophysiological aspect of identity. Specifically, Dfr has been found to be involved in the regulation of a specific immunoglobulin-containing cell-adhesion molecule called Beat-1c, and the *beat-ic* gene is likely a direct target of this transcription factor (Certel & Thor, 2004). Similarly, *eve* controls Unc-5, a receptor for the pathfinding guidance molecule netrin, in dorsally projecting motor neurons. Eve also controls the expression of genes involved in generating voltage- and ligand-gated currents in dorsally projecting motor neurons (Pym, Southall, Mee, Brand, & Baines, 2006).

THE CEREBRAL CORTEX

The cerebral cortex of mammals is the neural structure that is important for conscious thought, perception, and associative learning, as well as providing executive control over voluntary movement. The overall architecture of the

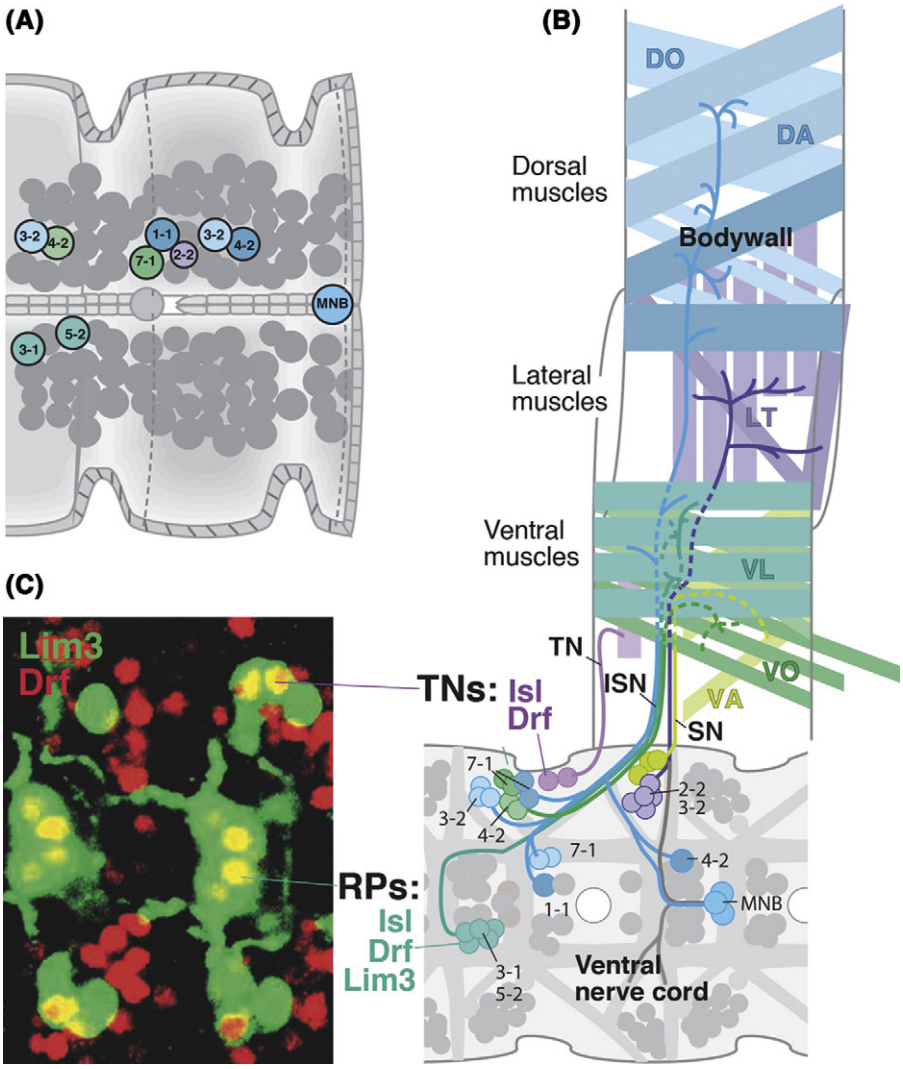


FIGURE 14.15 Specification of motor neurons in the *Drosophila* ventral cord. (A) Neuroblast map of two embryonic segments, showing in color the subset of neuroblasts that give rise to motor neurons. (B) Schematic flattened view of one body segment, showing in different colors the pattern of bodywall muscles and motor neurons innervating these muscles. The specific motor neuron(s) innervating each of the 30 muscle fibers comprising the musculature has been mapped. Motor neurons are color coded such that their color matches the one of the neuroblast from which it descends (see A) and the muscle it innervates. As in vertebrates, distinct groups of motor neurons are specified by expressing transcription factors of the LIM and POU family. Shown here and, photographically, in panel C are the RP neurons, expressing the combination Isl, Drf, and Lim3, and the transverse nerve (TN) motor neurons that express only Isl and Drf. Mutations in any of these genes result in motor neuronal abnormalities. *B from Certel and Thor (2004).*

cortex is relatively simple both in terms of its areal structure and the general composition of cells that make up the cortex. The neocortex, which forms most of the cerebral cortex in humans, is a six-layered structure, with each layer having its own particular array of neuronal cell types. Differences in the thicknesses and cellular densities of particular layers have enabled histologists to distinguish as many as 100 cytoarchitectonically distinct cortical areas in humans. There are two primary classes of neurons in the cortex. The majority of these cells, approximately 80%, are glutamatergic pyramidal neurons that originate from neuroepithelial cells in the ventricular zone of the telencephalic vesicles. Asymmetric divisions of these cells generate one daughter that remains neuroepithelial and apical, and another daughter that migrates basally. In most instances, the migrating daughter becomes a postmitotic cortical neuron, but as development proceeds, an increasing number of migrating daughter cells travel only a short distance to the subventricular zone, where they divide symmetrically to generate two neurons. The second major class of neurons in the cortex, GABAergic inhibitory interneurons, arise from the ventral telencephalon or ganglionic eminence (Anderson, Eisenstat, Shi & Rubenstein, 1997).

Cortical precursors, once generated, migrate away from the ventricular and subventricular zones, through an intermediate zone that will later be filled with axons, until they arrive at the cortical plate where they stop migrating and begin to differentiate into an enormous variety of specialized cortical cell types. This basal migration is coupled with an inside-out pattern of histogenesis in which the firstborn cells take up the deep layers of the cortical plate and the later-born cells migrate past the earlier generated neurons to occupy more superficial layers (Fig. 14.16A) (Angevine & Sidman, 1961). Cortical cells are specified with respect to their laminar fates before they migrate, for if postmitotic cells from the ventricular zone of a young embryo (destined for deep layers) are transplanted into the ventricular zone of an older one that is generating superficial cells, then the transplanted cells will

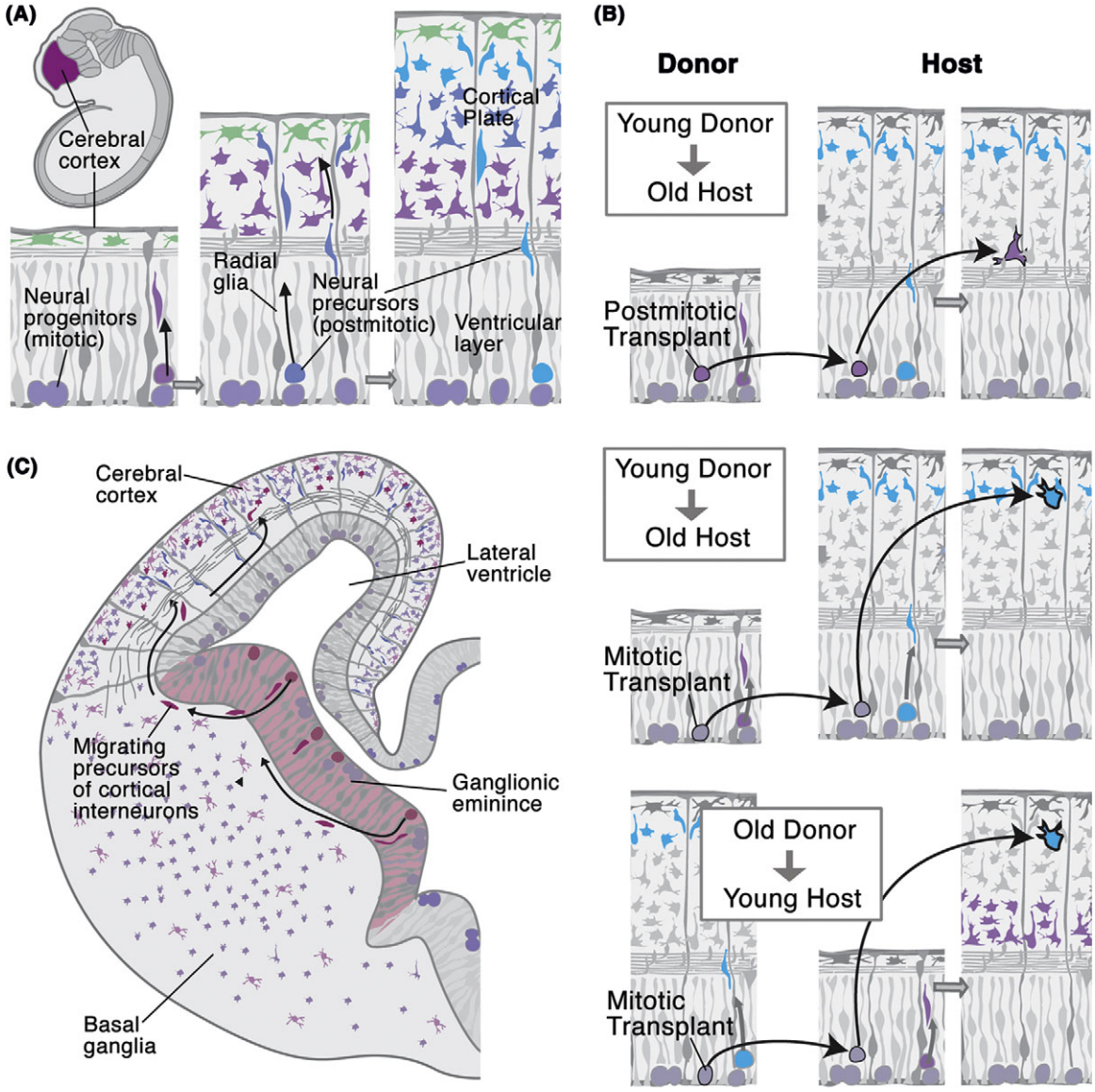


FIGURE 14.16 Laminar fate determination in the cerebral cortex. (A) Morphogenesis of the mammalian cerebral cortex. Neural precursors are born in the ventricular layer and migrate away from the ventricular surface, following tracks provided by radial glial cells. The firstborn cells are the Cajal-Retzius neurons (left in figure). Later-born neurons accumulate in a dense matrix of cells, the cortical plate (middle of figure). In this plate, neurons are ordered by birth date in such a way that older neurons (magenta) remain in deep layers, and younger neurons (blue) migrate through the deep layers to attain a superficial position (right in figure). (B) Top: If postmitotic ventricular cells from young donors (which normally would become deep cells) are transplanted into an old host, they do not adapt to their new environment and develop as deep layer neurons (arrow). However, if ventricular cells that have not yet exited the cell cycle are transplanted from young donors into an old host (middle panel), they adapt to their new environment and develop as superficial neurons (arrow). If, on the other hand, premitotic progenitors from old donors are transplanted into young brains, they can take later fates and become upper layer neurons. (C) Migration route of GABAergic interneurons from the ganglionic eminences in the ventral telencephalon into the cortical plate in a mammalian embryo.

still migrate to deep layers and have deep-layer fates (Fig. 14.16B, top panel). However, if progenitors from a young ventricular zone are in their last mitotic cycle when they are transplanted to the older environment, they can change their laminar identity, suggesting that cortical cell fate is still undetermined until the cells exit the cell cycle (Fig. 14.16B, middle panel). Interestingly, the reverse experiment of transplanting premitotic progenitors of old donors to younger ventricular zones does not result in these cells assuming earlier, deeper-layer fates (Fig. 14.16B, bottom panel), suggesting that progression of determination is a one-way street (McConnell, 1995).

The specification of excitatory neuron cell types appears to be largely under the control of intrinsic determinants, many of which are transcription factors. Pax6, Emx1,

FoxG1, and Lhx2 function at early times to regulate arealization and to broadly specify cortical progenitors. Cell type specification is closely linked to the laminar position, and while a number of genes have been identified that mark neurons in different laminae, in most instances the programs that specify neuronal cell identity are still poorly defined. Nonetheless, some key components of the developmental switch that controls the differentiation of subcortically versus intracortically projecting pyramidal neurons have been identified (Leone, Srinivasan, Chen, Alcamo, & McConnell, 2008). Neurons in laminar V project subcortically to the superior colliculus, hindbrain, and spinal cord, and they are derived from cells that express the Fezf2 transcription factor. Fezf2 in turns activates expression of the Ctip2 zinc-finger factor, while repressing expression of Satb2, a chromatin remodeling transcription factor. In the absence of Fezf2, glutamatergic pyramidal neurons, neurons in lamina V do not express Ctip2 and instead of sending their axons subcortically, they instead extend axons across the corpus callosum to other cortical areas. Interestingly, Satb2, is markedly upregulated in Fezf2 mutants. Satb2 is required for neurons in the cortex to cross the corpus callosum. Overexpression of Satb2 represses Ctip2 expression and results in fewer lamina V cells sending axons subcortically to the hindbrain and spinal cord. Conversely, overexpression of Fezf2 generates an excess of subcortically projecting neurons.

As noted previously, GABAergic neurons arise from the medial and caudal ganglionic eminences (MGE and CGE). They then migrate tangentially to the cortex, where they intercalate into the developing cortical plate and form connections with glutamatergic pyramidal neurons. Multiple inhibitory interneuron cell types are derived from these two domains (Anderson et al., 1997; Batista-Brito and Fishell, 2009). The MGE produces somatostatin-positive (SST^+) interneurons and fast spiking interneurons that express the calcium-binding protein parvalbumin (PV) and the potassium channel Kv3.1, which enable neurons to generate trains of high-frequency action potentials. While fast spiking PV neurons arise from ventral regions of the MGE, there is also a temporal gradient of PV versus SST interneuron differentiation, revealing spatial and temporal cues that regulate the development of these two cell types. Bipolar neuron cell types (calretinin and VIP cells) that are generated from the CGE develop later than their MGE-derived counterparts. Nkx2.1 is a key early determinant of GABAergic interneuron identity, by promoting MGE fates over LGE and CGE fates. This function may reflect an early role for Nkx2.1 in patterning the subpallium. Loss of Nkx2.1 results in a marked depletion of MGE derived neuronal cell types including PV^+ and SST^+ interneurons. Dlx1/2 and Mash1, which is regulated by Dlx1/2, are required for the specification of GABAergic cell types in the forebrain. These genes are global determinants of GABA neuron differentiation, as they are also required for the development of GABAergic neurons in the striatum, including medium spiny neurons.

Extrinsic signaling contributes to the laminar specification of cortical progenitors as indicated by experiments showing that BDNF is able to induce late-born progenitors to acquire early-born laminar fates. Extrinsic signals also appear to control the migration and distribution of inhibitory interneurons in the cortex. The loss of subcortically projecting pyramidal neurons in mice that lack Fezf2 is associated with changes in the laminar distribution of PV^+ and SST^+ interneurons, with reduced numbers of both cell types in lamina V (Lodato et al., 2011). Neuronal activity also regulates the development of subsets of inhibitory interneurons, including migration and their axonal and dendritic morphology. It therefore appears that neuronal differentiation in the cortex is regulated by time of birth, intrinsic transcriptional determinants (some of which may be activated in a time dependent manner—e.g., Fezf2 is expressed in deep-layer pyramidal neurons and is downregulated when upper-layer neurons that are born later are generated), and by extrinsic signals and activity as these cells migrate and intercalate into the cortical circuitry. Neuronal activity also plays a prominent role in regulating neuronal connectivity in the cortex. This is perhaps best exemplified by the activity-dependent refinement of the ocular dominance columns in the visual cortex (see Chapter 19).

CONCLUSIONS

The generation of neurons involves a variety of extrinsic signals and intrinsic transcription factors that confer individual identities on postmitotic neurons. Whether vertebrate or invertebrate, this process of determination involves a similar set of molecular and cellular mechanisms, such as regional and temporal identity genes, proneural genes, signaling molecules and receptors, the Notch pathway, and asymmetric cell division. All these conspire, over the course of neuronal generation, to create progenitors with distinct combinations of transcription factors that uniquely specify the fates of the differentiating neurons by regulating the expression of genes that affect specific features of the cell such as axon and dendrite growth and targeting, the topic of the next chapter.

References

Anderson, S. A., Eisenstat, D. D., Shi, L., & Rubenstein, J. L. (1997). Interneuron migration from basal forebrain to neocortex: Dependence on Dlx genes. *Science, 278,* 474–476.

Angevine, J. B. J., & Sidman, R. L. (1961). Autoradiographic study of cell migration during histogenesis of cerebral cortex in the mouse. *Nature, 192,* 766–768.

Appel, B., Korzh, V., Glasgow, E., Thor, S., Edlund, T., Dawid, I. B., & Eisen, J. S. (1995). Motoneuron fate specification revealed by patterned LIM homeobox gene expression in embryonic zebrafish. *Development, 121,* 4117–4125.

Batista-Brito, R., & Fishell, G. (2009). The developmental integration of cortical interneurons into a functional network. *Current Topics in Developmental Biology, 87,* 81–118.

Betschinger, J., & Knoblich, J. A. (2004). Dare to be different: Asymmetric cell division in *Drosophila*, C. elegans and vertebrates. *Current Biology, 14,* R674–R685.

Broihier, H. T., & Skeath, J. B. (2002). *Drosophila* homeodomain protein dHb9 directs neuronal fate via cross-repressive and cell-nonautonomous mechanisms. *Neuron, 35,* 39–50.

Bultje, R. S., Castaneda-Castellanos, D. R., Jan, L. Y., Jan, Y. N., Kriegstein, A. R., & Shi, S. H. (2009). Mammalian Par3 regulates progenitor cell asymmetric division via notch signaling in the developing neocortex. *Neuron, 63,* 189–202.

Campos-Ortega, J. A. (1995). Genetic mechanisms of early neurogenesis in *Drosophila* melanogaster. *Molecular Neurobiology, 10,* 75–89.

Cayouette, M., Poggi, L., & Harris, W. A. (2006). Lineage in the vertebrate retina. *Trends in Neurosciences, 29,* 563–570.

Certel, S. J., & Thor, S. (2004). Specification of *Drosophila* motoneuron identity by the combinatorial action of POU and LIM-HD factors. *Development, 131,* 5429–5439.

Cheesman, S. E., Layden, M. J., Von Ohlen, T., Doe, C. Q., & Eisen, J. S. (2004). Zebrafish and fly Nkx6 proteins have similar CNS expression patterns and regulate motoneuron formation. *Development, 131,* 5221–5232.

Cheng, L., Arata, A., Mizuguchi, R., Qian, Y., Karunaratne, A., Gray, P. A., … Ma, Q. (2004). Tlx3 and Tlx1 are post-mitotic selector genes determining glutamatergic over GABAergic cell fates. *Nature Neuroscience, 7,* 510–517.

Chenn, A., & McConnell, S. K. (1995). Cleavage orientation and the asymmetric inheritance of Notch1 immunoreactivity in mammalian neurogenesis. *Cell, 82,* 631–641.

Choksi, S. P., Southall, T. D., Bossing, T., Edoff, K., de Wit, E., Fischer, B. E., … Brand, A. H. (2006). Prospero acts as a binary switch between self-renewal and differentiation in *Drosophila* neural stem cells. *Developmental Cell, 11,* 775–789.

Dasen, J. S., Liu, J. P., & Jessell, T. M. (2003). Motor neuron columnar fate imposed by sequential phases of Hox-c activity. *Nature, 425,* 926–933.

Dasen, J. S., De Camilli, A., Wang, B., Tucker, P. W., & Jessell, T. M. (2008). Hox repertiores for motor neuron diversity and connectivity gated by a single accessory factor, FoxP1. *Cell, 134,* 304–316.

Del Barrio, M. G., Taveira-Marques, R., Muroyama, Y., Yuk, D. I., Li, S., Wines-Samuelson, M., … Richardson, W. D. (2007). A regulatory network involving Foxn4, Mash1 and delta-like 4/Notch1 generates V2a and V2b spinal interneurons from a common progenitor pool. *Development, 134,* 3427–3436.

Dorsky, R. I., Chang, W. S., Rapaport, D. H., & Harris, W. A. (1997). Regulation of neuronal diversity in the Xenopus retina by Delta signalling. *Nature, 385,* 67–70.

Elliott, J., Jolicoeur, C., Ramamurthy, V., & Cayouette, M. (2008). Ikaros confers early temporal competence to mouse retinal progenitor cells. *Neuron, 60,* 26–39.

Flores, G. V., Duan, H., Yan, H., Nagaraj, R., Fu, W., Zou, Y., … Banerjee, U. (2000). Combinatorial signaling in the specification of unique cell fates. *Cell, 103,* 75–85.

Gage, F. H. (2000). Mammalian neural stem cells. *Science, 287,* 1433–1438.

Garces, A., & Thor, S. (2006). Specification of *Drosophila* aCC motoneuron identity by a genetic cascade involving even-skipped, grain and zfh1. *Development, 133,* 1445–1455.

Glasgow, S. M., Henke, R. M., MacDonald, R. J., Wright, C. V., & Johnson, J. E. (2005). Ptf1a determines GABAergic over glutamatergic neuronal cell fate in the spinal cord dorsal horn. *Development, 132,* 5461–5469.

Goulding, M. D., Lumsden, A., & Gruss, P. (1993). Signals from the notochord and floor plate regulate the region-specific expression of two Pax genes in the developing spinal cord. *Development, 117,* 1001–1016.

Gowan, K., Helms, A. W., Hunsaker, T. L, Collisson, T, Ebert, P. J., Odom, R., & Johnson, J. E. (2001). Crossinhibitory activities of Ngn1 and Math1 allow specification of distinct dorsal interneurons. *Neuron, 31,* 219–232.

Harris, W. A., Stark, W. S., & Walker, J. A. (1976). Genetic dissection of the photoreceptor system in the compound eye of *Drosophila Melanogaster*. *Journal of Physiology, 256,* 415–439.

Isshiki, T., Pearson, B., Holbrook, S., & Doe, C. Q. (2001). *Drosophila* neuroblasts sequentially express transcription factors which specify the temporal identity of their neuronal progeny. *Cell, 106,* 511–521.

Jessell, T. M. (2000). Neuronal specification in the spinal cord: inductive signals and transcriptional codes. *Nature Reviews. Genetics, 1,* 20–29.

Jones, B. W. (2005). Transcriptional control of glial cell development in *Drosophila*. *Developmental Biology, 278,* 265–273.

Jusuf, P. R., Almeida, A. D., Randlett, O., Joubin, K., Poggi, L., & Harris, W. A. (2011). Origin and determination of inhibitory cell lineages in the vertebrate retina. *Journal of Neuroscience, 31,* 2549–2562.

Kempermann, G. (2011). Seven principles in the regulation of adult neurogenesis. *The European Journal of Neuroscience, 33,* 1018–1024.

Kessaris, N., Pringle, N., & Richardson, W. D. (2001). Ventral neurogenesis and the neuron-glial switch. *Neuron, 31,* 677–680.

Kiernan, A. E., Cordes, R., Kopan, R., Gossler, A., & Gridley, T. (2005). The Notch ligands DLL1 and JAG2 act synergistically to regulate hair cell development in the mammalian inner ear. *Development, 132,* 4353–4362.

Kriegstein, A., & Alvarez-Buylla, A. (2009). The glial nature of embryonic and adult neural stem cells. *Annual Review of Neuroscience, 32,* 149–184.

Lai, E. C., & Orgogozo, V. (2004). A hidden program in *Drosophila* peripheral neurogenesis revealed: fundamental principles underlying sensory organ diversity. *Developmental Biology, 269,* 1–17.

Lamb, T. D., Collin, S. P., & Pugh, E. N., Jr. (2007). Evolution of the vertebrate eye: Opsins, photoreceptors, retina and eye cup. *Nature Reviews. Neuroscience, 8,* 960–976.

Landgraf, M., & Thor, S. (2006). Development and structure of motoneurons. *International Review of Neurobiology, 75,* 33–53.

Landgraf, M., Roy, S., Prokop, A., VijayRaghavan, K., & Bate, M. (1999). Even-skipped determines the dorsal growth of motor axons in *Drosophila*. *Neuron, 22,* 43–52.

Leone, D. P., Srinivasan, K., Chen, B., Alcamo, E., & McConnell, S. K. (2008). The determination of projection neuron identity in the developing cerebral cortex. *Current Opinion in Neurobiology, 18,* 28–35.

Livesey, F. J., & Cepko, C. L. (2001). Vertebrate neural cell-fate determination: Lessons from the retina. *Nature Reviews. Neuroscience, 2,* 109–118.

Lodato, S., Rouaux, C., Quast, K. B., Jantrachotechatchawan, C., Studer, M., Hensch, T. K., & Arlotta, P. (2011). Excitatory projection neuron subtypes control the distribution of local inhibitory interneurons in the cerebral cortex. *Neuron, 69,* 763–779.

Ma, C., Zhou, Y., Beachy, P. A., & Moses, K. (1993). The segment polarity gene hedgehog is required for progression of the morphogenetic furrow in the developing *Drosophila* eye. *Cell, 75,* 927–938.

Marchetto, M. C., Carromeu, C., Acab, A., Yu, D., Yeo, G. W., Mu, Y., … Muotri, A. R. (2010). A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. *Cell, 143,* 527–539.

McConnell, S. K. (1995). Constructing the cerebral cortex: Neurogenesis and fate determination. *Neuron, 15,* 761–768.

Mollereau, B., & Domingos, P. M. (2005). Photoreceptor differentiation in *Drosophila*: From immature neurons to functional photoreceptors. *Developmental Dynamics, 232,* 585–592.

Mizuguchi, R., Kriks, S., Cordes, R., Gossler, A., Ma, Q., & Goulding, M. (2006). Ascl1 and Gsh1/2 control inhibitory and excitatory cell fate in spinal sensory interneurons. *Nature Neuroscience, 9,* 770–778.

Muller, T., Anlag, K., Wildner, H., Britsch, S., Treier, M., & Birchmeier, C. (2005). The bHLH factor Olig3 coordinates the specification of dorsal neurons in the spinal cord. *Genes & Development, 19,* 733–743.

Ohnuma, S., & Harris. W. A. (2003). Neurogenesis and the cell cycle. *Neuron, 40*, 199–208.

Ohsawa, R., & Kageyama, R. (2008). Regulation of retinal cell fate specification by multiple transcription factors. *Brain Research, 1192,* 90–98.

Peng, C. Y., Yajima, H., Burns, C. E., Zon, L. I., Sisodia, S. S., Pfaff, S. L., & Sharma, K. (2007). Notch and MAML signaling drives Scl-dependent interneuron diversity in the spinal cord. *Neuron, 53,* 813–827.

Petersen, P. H., Zou, K., Hwang, J. K., Jan, Y. N., & Zhong, W. (2002). Progenitor cell maintenance requires numb and numblike during mouse neurogenesis. *Nature, 419,* 929–934.

Poggi, L., Vitorino, M., Masai, I., & Harris, W. A. (2005). Influences on neural lineage and mode of division in the zebrafish retina in vivo. *Journal of Cell Biology, 171,* 991–999.

Pym, E. C., Southall, T. D., Mee, C. J., Brand, A. H., & Baines, R. A. (2006). The homeobox transcription factor Even-skipped regulates acquisition of electrical properties in *Drosophila* neurons. *Neural Development, 1*, 3.

Qian, X., Shen, Q., Goderie, S. K., He, W., Capela, A., Davis, A. A., & Temple, S. (2000). Timing of CNS cell generation: A programmed sequence of neuron and glial cell production from isolated murine cortical stem cells. *Neuron, 28,* 69–80.

Reinke, R., & Zipursky, S. L. (1988). Cell–cell interaction in the *Drosophila* retina: The bride of sevenless gene is required in photoreceptor cell R8 for R7 cell development. *Cell, 55,* 321–330.

Rousso, D. L., Gaber, Z. B., Wellik, D., Morrisey, E. E., & Novitch, B. G. (2008). Coordinated actions of the forkhead protein Foxp1 and Hox proteins in the columnar organization of spinal motor neurons. *Neuron, 59,* 226–240.

Schlosser, G. (2006). Induction and specification of cranial placodes. *Developmental Biology, 294,* 303–351.

Skeath, J. B., & Carroll, S. B. (1992). Regulation of proneural gene expression and cell fate during neuroblast segregation in *Drosophila*. *Development, 114,* 939–946.

Stam, F. J., Hendricks, T. J., Zhang, J., Geiman, E. J., Francius, C., Labosky, PA, … Goulding, M. (2012). Renshaw cell interneuron specialization is controlled by a temporally restricted transcription factor program. *Development*, [Epub ahead of print].

Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell, 126,* 663–76.

Takebayashi, S., Yamamoto, N., Yabe, D., Fukuda, H., Kojima, K., Ito, J., & Honjo, T. (2007). Multiple roles of Notch signaling in cochlear development. *Developmental Biology, 307,* 165–178.

Vallier, L., & Pedersen, R. A. (2005). Human embryonic stem cells: an in vitro model to study mechanisms controlling pluripotency in early mammalian development. *Stem Cell Reviews, 1,* 119–130.

Vrieseling, E., & Arber, S. (2006). Target-induced transcriptional control of dendritic patterning and connectivity in motor neurons by the ETS gene Pea3. *Cell, 127,* 1439–1452.

Waid, D. K., & McLoon, S. C. (1998). Ganglion cells influence the fate of dividing retinal cells in culture. *Development, 125,* 1059–1066.

Watanabe, T., & Raff, M. C. (1990). Rod photoreceptor development in vitro: Intrinsic properties of proliferating neuroepithelial cells change as development proceeds in the rat retina. *Neuron, 4,* 461–467.

Wichterle, H., Lieberam, I., Porter, J. A., & Jessell, T. M. (2002). Directed differentiation of embryonic stem cells into motor neurons. *Cell, 110,* 385–397.

Wodarz, A., & Huttner, W. B. (2003). Asymmetric cell division during neurogenesis in *Drosophila* and vertebrates. *Mechanisms of Development, 120,* 1297–309.

Zhou, Q., Choi, G., & Anderson, D. J. (2001). The bHLH transcription factor Olig2 promotes oligodendrocyte differentiation in collaboration with Nkx2.2. *Neuron, 31,* 791–807.

Zipursky, S. L., & Rubin, G. M. (1994). Determination of neuronal cell fate: Lessons from the R7 neuron of *Drosophila*. *Annual Review of Neuroscience, 17,* 373–397.

William A. Harris, Volker Hartenstein, and Martyn Goulding

This page intentionally left blank

CHAPTER

15

Neurogenesis and Migration

INTRODUCTION

The epithelium lining the brain vesicles in the neural tube, which is termed the ventricular zone (VZ), is the site of primary neurogenesis in the CNS. The idea that CNS neurogenesis occurs in the thin epithelium of the embryonic brain vesicles became apparent when Fred and Mary Sauer showed that cells in the neuroepithelium contain different amounts of DNA, with diploid cells located on the ventricular surface (Sauer & Chittenden, 1959). Clear images of the pseudo-columnar character of the neuroepithelium of embryonic rat brain were achieved in ^{3}H-thymidine labeling studies of the neocortex (Berry & Rogers, 1965) and in serial scanning and electron micrographs, which showed movements of the nuclei during the principal phases of the cell cycle (Hinds & Ruffett, 1971). Cells in interphase extend processes across the thin wall of the epithelium, one cell deep at this stage, and divide at the ventricular surface. During the cell cycle, the nucleus undergoes interkinetic migration, moving from the apical surface of the ventricular zone (VZ) during G1, entering S phase as the nucleus reaches the top of the VZ, and moving back toward the ventricular surface during G2 to enter mitosis, or M phase (Sauer & Walker, 1959). A cardinal feature of the neuroepithelium is its upside-down orientation relative to epithelia of other tissues. This relates to the fact that the neuroepithelium forms a tube that orients the apical surface of neuroblasts in the epithelial sheet toward the lumen of the neural tube and their basal surface toward the pia.

The neurons of the peripheral nervous system (PNS) and CNS originate from the neural tube and the folds of the neural tube. The cells of the PNS are generated first, as they delaminate in a specialized region of the dorsal aspect of the neural tube called the neural crest. Neural crest neuroblasts migrate out into the developing nonneuronal tissues, where they will form the axon tracts of the periphery, as the progenitors of the CNS are beginning to establish the regions of the brain and spinal cord that will receive the sensory information from the PNS and establish brain circuitry. We will therefore discuss neurogenesis and migration in the PNS first, followed by patterns of neurogenesis and migration in the CNS.

DEVELOPMENT OF THE PERIPHERAL NERVOUS SYSTEM

The Neural Crest is a Migratory Cell Population That Forms Multiple Derivatives

The *neural crest* is a transient population of cells, so named because they arise on the "crest" of the closing neural tube. This cell population is unique to vertebrates and forms most of the PNS. Neural crest cells migrate extensively along characteristic pathways and give rise to diverse and numerous derivatives. All of the *dorsal root, sympathetic, parasympathetic,* and *enteric ganglia* are derived from neural crest cells. Furthermore, most *cranial sensory ganglia* receive a contribution from the neural crest, with the remaining cells derived from *ectodermal placodes.* In addition to forming neurons and glia of the PNS, neural crest cells form melanocytes, cranial cartilage, and adrenal chromaffin cells. This wide variety of cell types arises from precursors in the neural folds and neural tube that are multipotent and have stem cell properties (Bronner-Fraser & Fraser, 1988; Stemple & Anderson, 1993). See Chapter 17 for a discussion of stem cells.

The neural crest originates at the border between the neural plate and the nonneural ectoderm by an inductive interaction between these two tissues (Selleck & Bronner-Fraser, 1995) (see Chapter 13). This induction initiates during gastrulation (Basch, Bronner-Fraser, & Garcia-Castro, 2006) even though precursors with the potential to

© 2013 Elsevier Inc. All rights reserved.

form neural crest only become recognizable by expression of neural crest markers within the dorsal portion of the neural tube. This delay occurs as the result of epigenetic changes that allow transcription of neural crest marker genes at the proper time (Strobl-Mazzulla, Sauka-Spengler, & Bronner-Fraser, 2010). These premigratory neural crest cells subsequently emerge from the neural tube and embark upon defined pathways.

Cells in the neural tube are epithelial and look much like contiguous soda cans that have a defined top (apical) and bottom (basal) side. They are closely apposed to one another and are connected by various types of adhesive junctions. In contrast, migratory cells such as neural crest cells are mesenchymal, having a fibroblast-like morphology that facilitates their movement. Thus, precursor cells within the neural tube change from an epithelial to a mesenchymal morphology as they turn into migratory neural crest cells. Such an *epithelial-to-mesenchymal transition* is a common event in development during the formation of tissues and organs. The transcription factor Snail2 is expressed in premigratory and early migrating cells and represents an early neural crest marker. Snail transcription factors have been associated with epithelial–mesenchymal transitions in a number of cell types, and Snail2 function is necessary for the emigration of neural crest cells. Initiation of neural crest cell migration proceeds in a rostral-to-caudal progression along most of the neural axis, following upon the heels of the head-to-tailward closure of the neural tube. After emigration, these cells move in a highly patterned fashion through neighboring tissues and localize in diverse sites.

Initiation of Migration

As they change from epithelial to migratory mesenchymal cells, neural crest cells undergo changes in adhesive properties. While within the neuroepithelium, neural tube cells express high levels of the *cell adhesion molecules* N-cadherin and cadherin-6B. Migrating cells downregulate these cadherins during migration and upregulate cadherin-7. Upon coalescing into ganglia and ceasing migration, cadherin-7 is downregulated and N-cadherin are again upregulated (Nakagawa & Takeichi, 1998). This suggests that a shift in cell surface and adhesive properties may accompany the onset and cessation of migratory behavior. Neural crest cells also turn on RhoB during the initiation of migration.

After leaving the neural tube, neural crest cells encounter extracellular spaces that are rich in *extracellular matrix* (ECM) *molecules*, such as fibronectin, laminin, collagens, and proteoglycans. These may serve as a good migratory substrate, and, indeed, the neural crest cell surface has abundant integrin receptors that mediate adhesion to ECM molecules. Furthermore, administration of function-blocking antibodies to the *β subunit of integrin* cause severe perturbations in neural crest development in the head. They appear to prevent the migration of some neural crest cells from the cranial neural tube. In the trunk, function-blocking antibodies that interfere with the *a_4 subunit of integrin* cause defects in neural crest cell movement but fail to alter the segmental pattern of migration through the somites. These results suggest that perturbing integrin function alters the properties of migratory cells, but not their overall metameric pattern of migration. Thus, cell–matrix interactions may play a permissive role in the migration of neural crest cells through the somites but cannot play an instructive role in directing the precise patterns or pathways for crest cell migration.

Techniques for Following Neural Crest Cells

When neural crest cells emerge from the neural tube, they first enter a cell-free space in which they are easily identifiable (Fig. 15.1). Subsequently, they invade other tissues in which they are difficult to distinguish. Therefore, it is necessary to *mark* neural crest cells in order to study their migratory patterns and derivatives. A variety of techniques have emerged for this purpose, ranging from transplantation of tissue containing premigratory neural crest cells to lineage tracers and molecular markers. *Neural tube transplantations* have provided a wealth of information about neural crest migratory pathways and, in particular, derivatives arising from this population. Although this approach was widely used in amphibians for studying numerous embryonic processes, it has been applied most successfully to the analysis of neural crest migratory pathways and derivatives in avian embryos (LeDouarin & Kalcheim, 1999). Initial experiments in birds involved transplanting a neural tube from a donor labeled with the radioactive marker [^{3}H] thymidine into an unlabeled host (Weston, 1963). Neural crest cells generated from the labeled neural tube were also labeled and could be

FIGURE 15.1 A transverse section through a chick embryo showing neural crest cells initiating migration from the dorsal neural tube and into an ECM-filled space. *Courtesy of Jan Lofberg.*

identified readily in the periphery. This technique yielded important information about early stages of neural crest migration, but the label became diluted with further cell division and was not useful for looking at long-term differentiation of neural crest cells into diverse derivatives.

To circumvent this problem, LeDouarin took advantage of the ability to perform grafts between related species of birds, such that the grafted cells were indelibly marked (Fig. 15.2). This created an interspecific *chimera*, from the Greek meaning "fabulous monster," containing a donor quail portion of the neural tube and neural crest in an otherwise normal chick host embryo. Quail neural crest cells migrate away from the grafted neural tubes and can be recognized easily within the host chick embryo by staining for condensed heterochromatin, which characterizes the quail but not chick cells. This technique has been facilitated greatly by the advent of quail-specific antibodies. Recently, the availability of transgenic chickens with cell populations that are labeled with green fluorescent protein (GFP) makes it possible to perform similar grafts within the same species. Use of the these transplantation studies has made it possible to demonstrate that neural crest cell populations originating from different axial levels follow distinct migratory pathways and give rise to different progeny once they reach their destinations.

In addition to grafting experiments, antibodies that recognize neural crest cells, such as HNK-1 and NC-1 antibodies, made it possible to identify early migrating neural crest cells without the necessity of performing microsurgery. The exclusive use of neural crest antibodies has several pitfalls because these antibodies are neither entirely specific nor stain the full complement of neural crest cells. However, they do provide important confirmatory information about the pathways followed by neural crest cells. More recently, a number of other molecular markers for early neural crest populations have become available. These include the cell adhesion molecule cadherin-6B and the transcription factor Snail2, expressed in the neural folds and neural crest cells during early stages of migration, and Sox10, expressed in premigratory, early migrating neural crest cells and later in neural crest-derived glia.

One problem with using antibodies or molecular markers to follow cell migratory patterns is that these do not represent true lineage markers. Shifts in their expression patterns could just as easily reflect up- or downregulation of these molecules as changes in cell position. Therefore, approaches for labeling small groups, as well as large populations, of neural crest cells have been employed to follow individual cell movements and interactions within the population. One successful approach has been to inject the *lipophilic vital dye DiI* into the neural tube or neural folds. Because the dye is hydrophobic and lipophilic, it intercalates into all cell membranes that it contacts. Injection into the neural tube marks all neural tube cells, including presumptive neural crest cells, within its dorsal aspect. Because the time and location of injection can be controlled, this provides a direct approach for following migratory pathways. When dye injections are made focally into neural folds, this technique can be used to label very small numbers of cells. It has the further advantage of being applicable to almost all vertebrates. Similarly, the availability of green fluorescent protein (GFP)

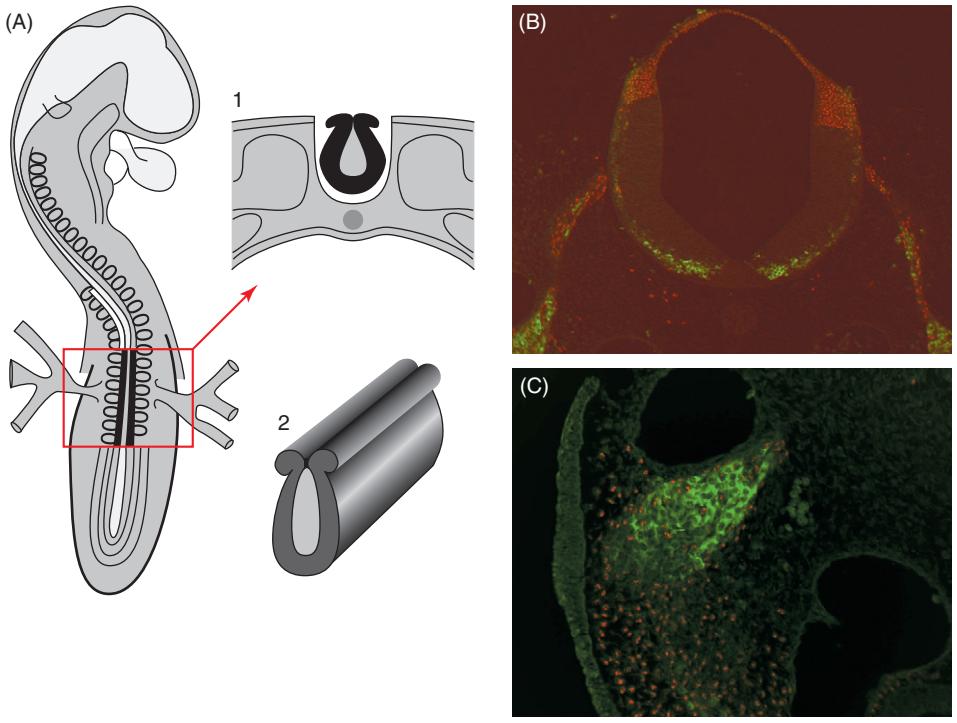


FIGURE 15.2 Procedures for grafting a fragment of the neural primordium from a donor quail into a host chicken embryo as used by LeDouarin and colleagues. (A) View of an avian embryo with anterior at the top. Neural folds are shown in black in the boxed region; this structure is removed and transplanted to a host embryo. (1) Cross-section through the embryo in regions shown in the box with the neural tube (2) shown in black. From LeDouarin (1982). (B) An example of a section through an embryo after grafting of a quail neural tube into a chick host. Quail cells are recognized by a quail-specific antibody (red), whereas neurons are marked in green with a neurofilament marker. (C) A higher magnification section showing quail cells (red nuclei) incorporated into a neural crest-derived ganglion, stained green with a neurofilament antibody. *Courtesy of Anne Knecht and Clare Baker.*

and its variants has made it possible to label neural tube cells, including premigratory neural crest cells, by electroporating constructs encoding GFP into the neural tube lumen at various stages and axial levels. A disadvantage of these approaches, however, is that the dye or fluorophore becomes diluted with each cell division and, therefore, this approach is not useful for examining the long-term differentiation of neural crest derivatives.

Because most studies of cell migration *in vivo* look only at the beginning and end point, little is known about the dynamics of neural crest cell movement. Advances in imaging technologies have made it possible to acquire more refined images *in ovo*. Time-lapse movies of cranial neural crest migration can be generated by injecting DiI or electroporating GFP constructs (Yasuda et al., 2000) into the lumen of the chick neural tube, and high-resolution confocal images can be made of the labeled cells over time (Fig. 15.3). These movies make it possible to visualize the migratory behavior of neural crest cells as they emerge from the cranial neural tube and migrate toward the branchial arches. As the embryo develops, it is possible to follow individual cell movements within the branchial arches, which will give rise to the bone and cartilage of the jaw. Neural crest cells move in defined streams and appear to form distinct chains in which cells are oriented along the same trajectory and remain in contact via their processes. This suggests a large degree of intercommunication between migrating populations of cranial neural crest cells.

A number of different methods have proved useful for following the pathways of neural crest migration, including neural tube transplantations, vital dye labeling, GFP expression, and antibody staining. These have made it

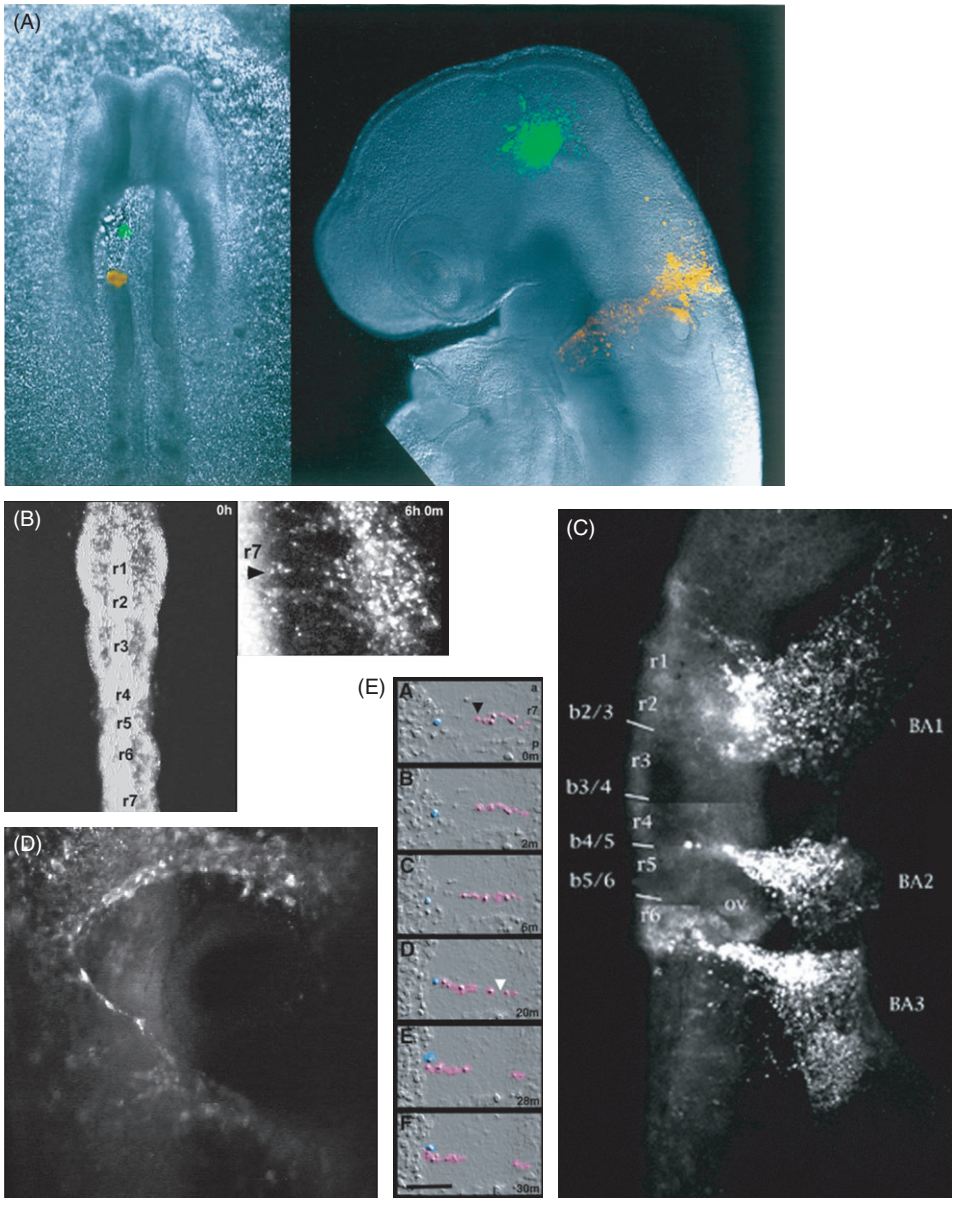


FIGURE 15.3 Labeling of neural crest cells with liphophilic tracer dyes. (A) On the left is a chick embryo viewed shortly after unilateral ablation of the neural folds of roughly half the neural tube at the level of the presumptive midbrain. A spot of DiO (green) was focally injected into the remaining neural tube, and DiI (red) was injected into the bordering intact neural folds. On the right, the same embryo after 48 h of further development. Cells that were originally in the neural tube had dispersed as migratory neural crest cells from both the green- and the red-labeled spots. (B–D) Different views of time-lapse movies of embryos in which the neural tube and premigratory neural crest cells were labeled with DiI. (B) Two views of the same embryo immediately after DiI labeling (left) and several hours after neural crest migration from the hindbrain. Rhombomeres (r1–r7) are indicated. (C) A similar embryo at a later time point by which time neural crest cells have migrated into the branchial arches (BA). (D) At higher magnification of neural crest streams in the branchial arches, cells seem to be in close contact as if following each other in narrow streams. (E) Following individual cells by time-lapse cinematography demonstrates close and maintained connections between individual cell pairs. *Courtesy of Paul Kulesa and Scott Fraser.*

possible not only to establish the various routes of neural crest migration occurring at different axial levels but also to follow the derivatives of neural crest cells arising from distinct locations.

Regionalization of the Neural Crest along the Body Axis

Despite differences in the nature of the techniques involved, transplantation experiments, molecular markers, and direct labeling have provided similar pictures of the migratory pathways followed by neural crest cells. These migratory pathways and the derivatives formed by neural crest cells are regionalized according to their original position along the anterior/posterior axis, such that cells from a given axial level give rise to a characteristic array of progeny and follow distinct pathways from those arising at other axial levels (LeDouarin & Kalcheim, 1999; Noden, 1975). The different populations of neural crest cells arising along the neural axis have been designated as cranial, vagal, trunk, and lumbosacral (Fig. 15.4). Distinct cell types differentiate from these different populations.

At cranial levels, some neural crest cells contribute to the *cranial sensory ganglia* and the *parasympathetic ciliary ganglion* of the eye, whereas others migrate ventrally to form many of the *cartilaginous elements of the facial skeleton*. One of the interesting things about neural crest-derived bones of the head is that these are the only skeletal elements in the body that are derived from ectoderm. Precise quail/chick grafting experiments have determined the regions of neural tube from which neural crest cells arise to contribute to cartilaginous elements and cranial ganglia (Noden, 1975). Neural crest cells originating in the midbrain migrate primarily as a broad, unsegmented sheet under the ectoderm; they contribute to derivatives ranging from the skeleton around the eye, connective tissue and membranous bones of the face, to the ciliary ganglion and trigeminal ganglia (LeDouarin & Kalcheim, 1999). Neural crest cells arising in the hindbrain migrate ventrally and enter the branchial arches to form the bones of the jaw.

Vagal neural crest cells migrate long distances to form the *enteric nervous system*, which also receives a contribution from the lumbosacral neural crest. Within the gut, the earliest generated crest cells move as a wave from anterior to posterior to populate the bowel, which they appear to populate in sequence such that the anterior portions are populated by neural crest cells first and the cells move to progressively more posterior sites. Mice bearing a "lethal spotted" mutation lack neural crest cells in a portion of their bowel. This leads to a lack of innervation in the region called aganglionic bowel. As a consequence, food waste fails to move through this portion of the bowel, leading to a disorder called "mega-colon." A similar defect is observed in humans leading to "Hirschsprung's disease." This disorder arises from a defect in molecules

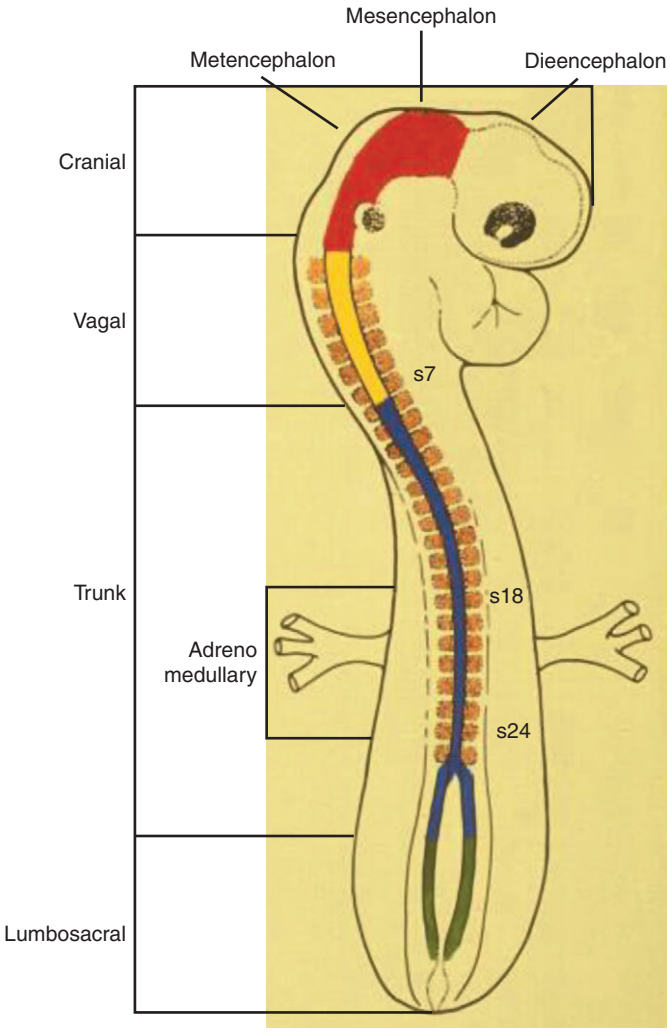


FIGURE 15.4 Schematic diagram showing different levels of the neural axis from which neural crest cells arise. From anterior to posterior, the neural axis can be divided into cranial, vagal, trunk, and lumbosacral levels. Each gives rise to distinct derivatives. *Drawing provided by Dr. M. Bronner.*

called endothelins and their receptors. The failure of neural crest migration in the aganglionic bowel of lethal spotted mutant mice is caused by a defect in mesenchymal components of the gut, whereas the neural crest cells themselves are normal.

Trunk neural crest cells follow two primary migratory pathways (Fig. 15.5): a *dorsolateral pathway* between the ectoderm and the somite and a *ventral pathway* through the rostral half of each sclerotome, the mesenchymal portion of the somite, which will go on to form the vertebrae. Cells following the dorsolateral stream give rise to *melanocytes*. Those cells following the ventral pathway give rise to the PNS of the trunk, including the chain of *sympathetic ganglia* and *dorsal root ganglia*, as well as *chromaffin cells* of the adrenal medulla. In addition to these neurons, these cells generate *glia* of the peripheral ganglia and *Schwann cells* that ensheathe and myelinate peripheral axons.

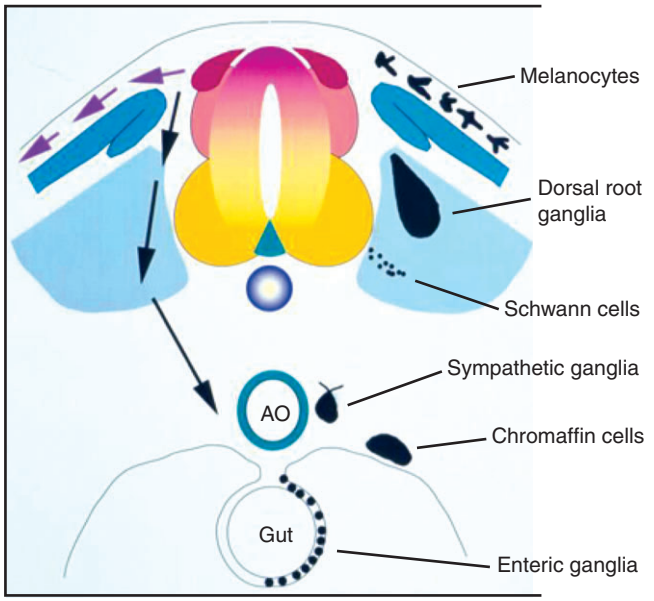


FIGURE 15.5 Schematic diagram of an idealized embryo in cross section showing pathways of neural crest migration in trunk and derivatives formed. Neural crest cells migrate along two primary pathways: dorsally under the skin or ventrally through the sclerotome. Dorsal migrating cells form pigment cells, whereas ventrally migrating cells give rise to dorsal root and sympathetic ganglia, Schwann cells, and cells of the adrenal medulla. *Drawn by Mark Selleck.*

The neural crest can be subdivided into specific subpopulations based upon their level of origin along the neural tube. Different populations follow different migratory pathways and form characteristic types of derivatives. For example, cranial bone and cartilage only arise from the cranial neural crest.

Grafting of Neural Crest Cells to New Locations—Intrinsic Versus Extrinsic Cues?

Neural crest cells that arise and migrate at different axial levels assume different fates. Thus, it is possible that they are specified at early times to take on particular fates. For example, one possibility is that neural crest cells are preprogrammed by their axial level to populate specific derivatives; alternatively, neural crest cells might be multipotent and migrate naively into available locations, where local cues provide them with instructions about their fates. To test the role of migratory pathways in choice of fate, neural tubes from particular axial levels have been grafted to new locations. This is referred to as a "heterotopic" graft. These experiments have revealed, for example, that when vagal neural crest cells are grafted to trunk regions, they form normal trunk derivatives (dorsal root ganglia, sympathetic ganglia, etc.) and normal vagal derivatives (enteric ganglia of the gut) (Fig. 15.6). The gut is immediately ventral to the trunk region and is connected to it by a narrow piece of tissue called the dorsal mesentery. Despite their close proximity, trunk neural crest cells

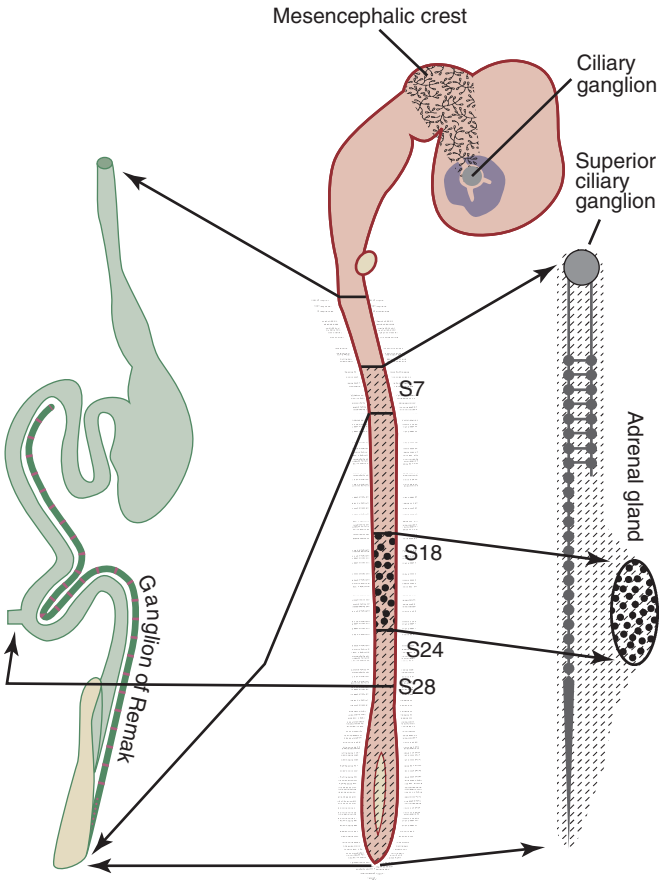


FIGURE 15.6 Neural crest cells at different levels along the anterior–posterior axis give rise to distinct autonomic and adrenomedullary derivatives in avian embryos. In the cephalic region (center), mesencephalic crest cells populate the ciliary ganglion. Ganglia of the sympathetic chain (right), including the superior ciliary ganglion, are formed from spinal neural crest cells originating caudal to somite 5. Cells of the adrenal medulla (right) originate exclusively from neural crest cells between somites 18 and 24. Vagal neural crest cells generated between somites 1 and 7 form enteric ganglia (left), whereas cells of the ganglion of Remak (left) are derived from the lumbosacral neural crest posterior to somite 28. *Drawing provided by Dr. M. Bronner.*

normally fail to invade the gut where enteric ganglia form. When vagal neural tubes are grafted in place of trunk neural tubes, however, donor vagal neural crest cells do invade the gut and form enteric ganglia. Therefore, these cells can respond to normal trunk neural crest migratory pathways, but, in addition, some cells behave in a "uniquely vagal" fashion and migrate directionally to the gut. This indicates some intrinsic differences between the two populations. Along similar lines, when cranial neural tubes are grafted in place of the trunk neural tube, donor cranial neural crest cells form some normal derivatives in the trunk-like dorsal root and sympathetic ganglia. Other cells, however, fail to migrate and instead differentiate into ectopic cartilage. In the reciprocal experiment, avian trunk neural crest cells grafted to the head region appear unable to generate cartilage at all, although they

can participate in the formation of elements of the cranial ganglia (LeDouarin & Kalcheim, 1999) (Fig. 15.6).

These experiments suggest two things: (1) environmental factors can influence neural crest cells from different axial levels to express a broader range of fates than they would normally express when left *in situ*, and (2) some neural crest cells undergo their intrinsic program even when grafted to an ectopic site. Thus, some combination of *intrinsic and extrinsic information* is likely to govern neural crest cell fate decisions. The migratory pathways taken by neural crest cells are likely to play an important regulatory role in cell fate specification. Indeed, different migratory pathways contain different distribution patterns of important inducing factors. A number of factors, including BMPs, neuregulins, and glucocorticoids, have been demonstrated to influence the choice of neural crest cells into neural, glial, or chromaffin lineages (see Chapter 18). When added to multipotent neural crest stem cells in culture, these factors can drive these cells into neural, glial, or chromaffin lineages, respectively. Accordingly, BMP-7 is present in the dorsal aorta adjacent to the location where trunk neural crest cells differentiate into sympathetic neurons. Similarly, glucocorticoids are produced by the adrenal cortex, which surrounds the adrenal medullary cells.

In addition to specific growth factors, the timing of emigration may play an important role in eventual neural crest cell fate decisions. Neural crest cells exhibit an orderly pattern in the timing of migration, with cells initially following the ventral pathway and later contributing to progressively more dorsal derivatives. Although early and late migrating cranial neural crest cells appear to have a similar developmental potential, it is likely that the cell population does undergo a change in the range of fates that the cells can assume. For example, the last emigrating trunk neural crest cells in the bird give rise to pigment cells and appear to have a limited capacity to form sympathetic neurons. However, *stem cells* appear to persist in the neural tube and retain the potential to form neural crest. Well past the normal time of neural crest cell emigration, cells with the full range of neural crest potential can be isolated from the ventricular zone of the developing spinal cord (Sharma, Korade, & Frank, 1995). In the embryo, these late-emigrating cells appear to contribute to a subpopulation of neural crest derived cells in the dorsal root ganglia.

Summary

Neural crest cells are somewhat plastic with respect to their prospective fates. When put into a new environment, they sometimes behave according to their new location. However, some cells act according to their original location and thus have some "intrinsic" information. With time, the developmental potential of neural crest populations becomes restricted, although a subpopulation of "stem cells" may remain until late stages.

SEGMENTAL MIGRATION OF NEURAL CREST CELLS

A hallmark of the developing peripheral nervous system is its inherent *segmentation.* After neural crest cells migrate from the neural tube and through the somites, they condense to form segmentally arranged sensory and sympathetic ganglia. For each somite, a single sensory and sympathetic ganglion forms. This exquisite and reproducible pattern suggests the presence of some inherent segmental information in the embryo that is responsible for segmental migration and gangliogenesis of neural crest cells. In longitudinal sections through early embryos, it is clear that neural crest cells migrate in a metameric pattern, moving exclusively through the rostral half of each somite while failing to enter the caudal half (Fig. 15.7).

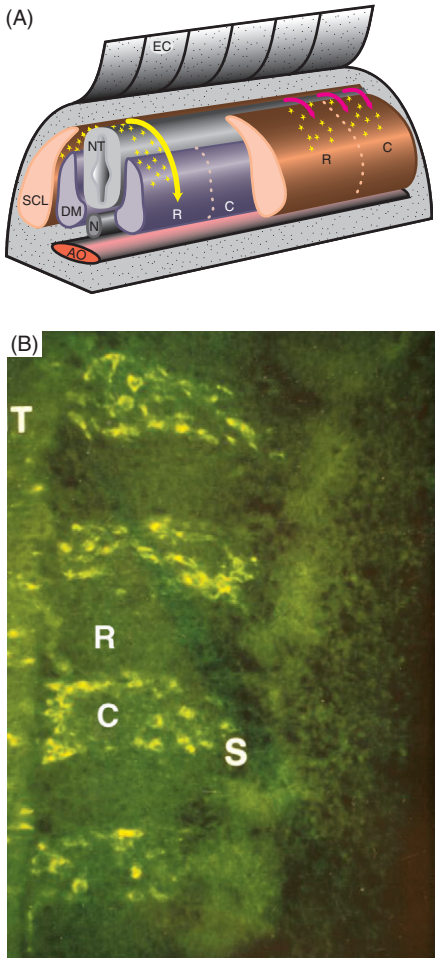


FIGURE 15.7 Trunk neural crest cells migrate in a segmental fashion. (A) Schematic diagram demonstrating that neural crest cells migrate through the sclerotomal portion of the somites, but only through the rostral half of the sclerotome. (B) In longitudinal section, neural crest cells (green) can be seen migrating selectively through the rostral half of each somitic sclerotome (S). *From Bronner-Fraser (1986).*

Such a segmental pattern of migration could be caused by inherent cues within the neural tube that direct neural crest cells to migrate segmentally. Alternatively, tissues through which neural crest cells migrate may contain patterning information that results in segmental migration. The relationship between neural crest cells and their surrounding tissues has been explored by manipulating the neural tube and/or somites in a series of grafting experiments. Removal of the somites results in the formation of huge, unsegmented neural crest–derived ganglia, suggesting that somites are necessary for the segmental migration of neural crest cells. However, inversion of the segmental plate in the rostrocaudal dimension reverses the pattern of neural crest migration such that neural crest cells migrate through the half of the rotated sclerotome that was originally rostral but is now caudal. Motor axons, which also traverse the rostral sclerotome in normal animals, exhibit similar behavior to neural crest cells after segmental plate rotation. This suggests that the information necessary to guide neural crest cells and motor axons is intrinsic to the somites. Furthermore, these experiments show that the rostrocaudal polarity of the somites is already established at the segmental plate stage.

Other experimental manipulations demonstrate that the caudal half of each somite is inhibitory, whereas the rostral half somite is permissive for neural crest migration and motor axon guidance. If somites are constructed to contain only caudal sclerotome tissue, neural crest cells and motor axons fail to migrate altogether. Conversely, an all-rostral somite results in the absence of segmentation for both neural crest cells and motor axons. Taken together, these findings demonstrate that the segmental migration of both neural crest cells and motor axons is due to cues inherent in the somite. These cues may be caused by attractive cues in the rostral-half sclerotome, inhibitory cues in the caudal-half sclerotome, or a combination of both. The finding that somites containing all caudal-half sclerotomes cannot support neural crest migration strongly suggests that some of the guidance cues are inhibitory. This segmentally arranged stream of migrating neural crest cells in turn results in the segmental organization of the neural crest–derived ganglia of the PNS.

Although it was suggested that Eph/ephrin signaling might direct the pattern of trunk neural crest migration (Krull et al., 1997; Wang & Anderson, 1997), the Eph and ephrin mutant mice that have been examined fail to exhibit trunk neural crest migration defects (Oriolli et al., 1996; Wang, & Anderson, 1997). Another receptor expressed by neural crest cells is Neuropilin-2 (Npn2), which is expressed in premigratory and migratory neural crest in both chick and mouse. Strikingly, trunk neural crest cells no longer migrate segmentally through the somites of *Npn-2* mutants and are distributed equally in rostral and caudal sclerotome. These defects are the result of Npn2/Semaphorin3F (Sema3F) signaling, as *Sema3F* mutants

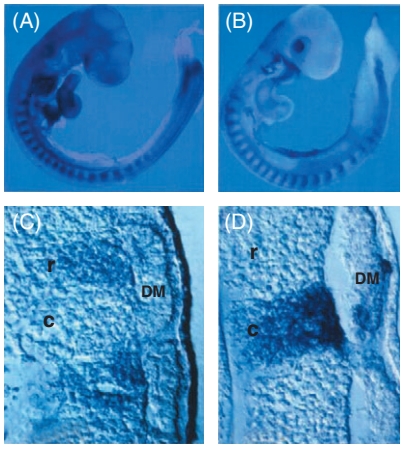


FIGURE 15.8 Distribution of Eph receptors (A and C) on neural crest cells in the rostral sclerotome and ephrin ligands in the caudal sclerotome (B and D) of chick embryos. Eph receptors are on neural crest cells in the rostral half of each somite, whereas inhibitory ephrin ligands are expressed in the caudal halves of each sclerotome. *From Krull et al. (1997).*

have an identical phenotype. These results demonstrate that Npn-2/Sema3F signaling is necessary to restrict neural crest migration to the rostral somite (Gammill, Gonzalez, Gu, & Bronner-Fraser, 2006). Interestingly, metameric dorsal root ganglia still form in Npn-2 mutants in the absence of segmental neural crest migration, suggesting that segmental migration may be separable from ganglion formation. However, mice lacking both *Npn1* and *Npn2* form abnormally fused ganglia, suggesting that Npn-1 is required for the later segregation of neural crest-derived ganglia (Fig. 15.8) (Gammill et al., 2006).

As in the trunk, the migration of neural crest cells in the hindbrain is also segmented. Three broad streams of migrating cells are found adjacent to rhombomere 2 (r2), r4, and r6, whereas no neural crest cells are apparent adjacent to r3 and r5. Focal injections of DiI at the levels of r3 and r5 have demonstrated that both of these rhombomeres generate neural crest cells and that the apparent segmental pattern results from the DiI-labeled cells that originated in r3 and r5 deviating rostrally or caudally and failing to enter the adjacent preotic mesoderm or otic vesicle region.

Trunk neural crest cells move in a segmental pattern, which is controlled by inhibitory molecules present in the caudal half of each somite. Neural crest cells have Npn2 receptors on their surface. When they encounter Sema3F ligands in the caudal somite, they are diverted from this location, leading to selective migration through only the rostral-half somite.

Cessation of Neural Crest Migration

Neural crest cells exhibit an orderly pattern in their migration. Whereas the first cells to migrate tend to move

most ventrally, later migratory cells contribute to progressively more dorsal derivatives. The different destinations of neural crest derivatives are therefore populated in a sequential order during development. One possibility is that the early migrating cells "fill" the more ventral sites, effectively clogging up the pathway. Alternatively, the sites themselves may change with time so that they can no longer support migration.

Surprisingly little is known about how neural crest cells know that they have reached the appropriate destination and stop their migration. This is in marked contrast to the developing CNS, where a number of mutations affect the ability of neuroblasts to cease migrating. It is clear that cell adhesion molecules such as N-cadherin and N-CAM are upregulated after cells reach their final sites and condense to form peripheral ganglia. However, there is no evidence for a causal role for cell adhesion molecules in this process.

Neurogenesis in the PNS

Some of the factors and signaling cascades that result in neurogenesis in the developing PNS are beginning to be understood (see Chapter 16). Many neural crest cells appear to be multipotent and not yet committed to a neural fate. However, particular external factors can influence their fate decisions. For example, clonal populations of neural crest cells become neural in the presence of BMPs and glial in the presence of glial growth factor. Similarly, activation of Notch signaling promotes gliogenesis at the expense of neurogenesis in neural crest cells and their derivatives. Certain transcription factors may bias cells toward certain fates; for example, while Mash-1 is essential for sympathetic neuron formation, Neurogenins are essential for sensory fates (Christiansen, Coles, & Wilkinson, 2000).

An interesting contrast between the developing PNS and the CNS is that migrating neural crest cells proliferate as they move. In fact, even after exhibiting defined neuronal characteristics, some neural crest derivatives continue to divide. For example, in the developing sympathetic ganglia, neural crest-derived cells express neurotransmitters and other proteins characteristic of sympathetic neurons but remain actively mitotic. However, other neural crest cells, such as sensory neurons, appear to withdraw from the cell cycle well before they express neuronal traits.

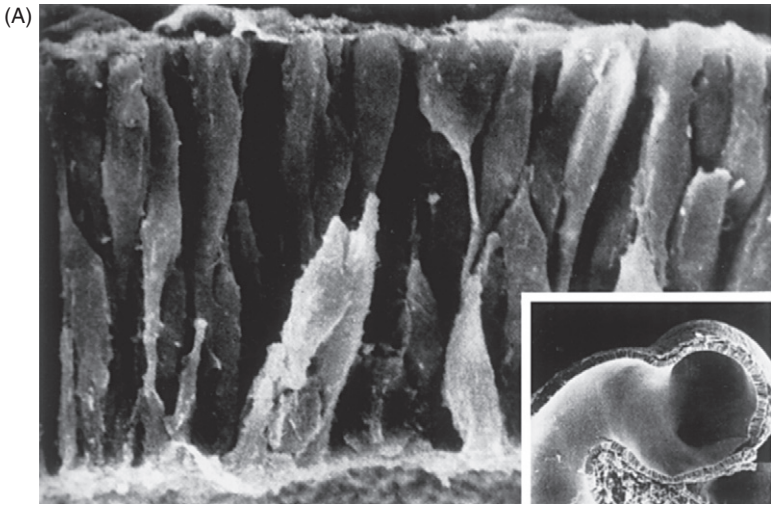
Summary

Neural crest cells migrate over long distances throughout the body to form diverse cell types, including neurons, glia, melanocytes, and cells of the adrenal gland. The migratory pathways of neural crest cells vary along the rostrocaudal body axis, and the local environments through which cells migrate and eventually differentiate play an important role in phenotypic specification. The migration of neural crest cells is guided by both positive (attractive or permissive) and inhibitory (repulsive) cues that are found in the environment.

CELL MIGRATION IN THE CNS

The emergence of the laminar architecture of the mammalian neocortex depends on neurogenesis in germinal zones, migration of postmitotic neurons (Caviness, 1982; Del Rio, Martinez, Auladell, & Soriano, 2000; McConnell, Ghosh, & Shatz, 1994; Wood, Martin, & Price, 1992; Zecevic & Rakic, 2001), formation of transverse zones of transient neurons (the preplate, cortical plate, marginal zone, and subplate) (Caviness, 1982; Del Rio et al., 2000; Marin-Padilla, 1998), and assembly of postmigratory cortical neurons into six neuronal layers. The columnar organization of cortical regions of the CNS develops from the radial glial scaffold, which provides a substrate for directed cell migrations of the primary output neurons of the neocortex in early periods of development, and generates neurons in later stages of development. Although the anatomists Retzius, Kolliker, and Magani described a system of radial glial cells in the nineteenth century, the influence of radial glial cells on cortical histogenesis did not become apparent until the late twentieth century. The spatiotemporal pattern of neurogenesis and migration during the formation of neural layers in developing cortex emerged from the pioneering work of Richard Sidman (1970) on thymidine "birthdating." The broad application of this approach to different brain regions and times of development demonstrated the spatiotemporal pattern of neurogenesis in the vertebrate CNS. The discovery of directed neuronal migration along glial fibers by Pasko Rakic in the 1970s (Rakic, 1972; Sidman & Rakic, 1973) set the stage for a period of intense focus on the process of neuronal migration in cortical histogenesis (Sidman & Rakic, 1973) (Fig. 15.10). Cell and molecular studies of identified neurons from the cerebellar cortex and hippocampal formation provided real-time imaging that defined the mode of movement of neurons migrating along glial fibers (Edmondson & Hatten, 1987; Gasser & Hatten, 1990a).

Studies on the architectonics of human cortical malformations (Barkovich et al., 1996; Dobyns & Truwit, 1995) revealed populations of heterotopic cells and suggested that pathogenic processes disrupted normal glial-guided neuronal migration (Rakic, 1988). The hypothesis that cortical malformations resulted from migration defects became even more attractive with the analysis of spontaneously occurring neurological mutations. All of these studies underscore the critical role of neuronal migration in brain development.



Academic press items and derived items
copyright © 1999 by academic press

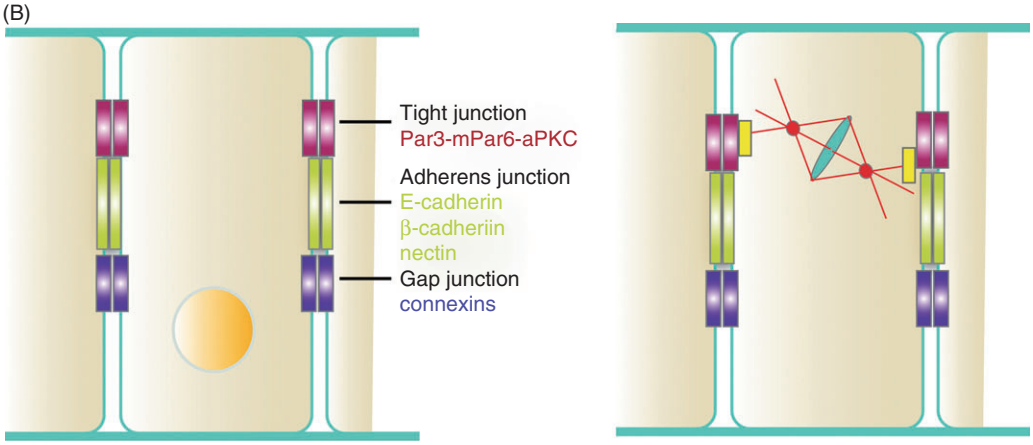


FIGURE 15.9 (A) *The Ventricular Zone Forms a Pseudostratified Columnar Epithelium*. Here neural progenitor cells have been visualized in the cerebral vesicle of a hamster embryo using scanning electron microscopy. Neuroepithelial cells are elongated bipolar cells that, at this early stage of development (E9.25), span the entire wall of the cerebrum. Some of the cells at the ventricular surface (bottom) appear spherical; these cells have retracted their cytoplasmic processes and are presumably rounding up in preparation for mitosis. Other rounded cells at the external surface (top) may be young neurons beginning to differentiate. (Inset) A low-power view of the hamster cerebral vesicle, corresponding roughly to that of a human embryo at the end of the first month of gestation.(B) *Polarity Proteins and Vertebrate Epithelia*. Three classes of cell-cell junctions form between cells in a vertebrate epithelium. Conserved polarity proteins function in the formation of tight junctions (red). The cadherin and nectin cell adhesion proteins are required to form adherens junctions (green) and connexins form gap junctions (purple). During cell division in the neuroepithelium, the mPar6 polarity complex positions the spindle of the dividing cell (see text for details). *From Sidman and Rakic (1973).*

Cortical Histogenesis: Formation of the Basic Embryonic Zones

During corticogenesis, early-born, polymorphic neurons settle in a narrow zone above the VZ, termed the preplate (PP) (Caviness, 1982; Del Rio et al., 2000; McConnell et al., 1994; Wood et al., 1992; Zecevic & Rakic, 2001) (Fig. 15.11). Preplate neurons include Cajal-Retzius cells, GABAergic interneurons, and future subplate projection neurons (Caviness, 1982; Del Rio et al., 2000; Marin-Padilla, 1998). During the preplate stage, preplate neurons pioneer the first axon projections of the neocortex, with some preplate neurons extending descending axons to subcortical targets and others extending corticocortical axons to targets in the ipsilateral or contralateral hemisphere via the anterior commissure. As the preplate matures, extensive cell movements occur across the dorso-ventral (DV) plane of the neuroaxis. At late preplate stages, preplate neurons undergo changes in polarity, shifting from a random orientation into a radial orientation (Schneider, Gulacsi, & Hatten, 2011). Thereafter, the preplate separates into a superficial marginal zone (MZ) of Cajal-Retzius cells and small neurons, the cortical plate (CP) (Rakic & Zecevic, 2003), and a deeper layer called the subplate (SP) (Fig. 15.11).

Studies on the genetics of cortical development underscore the importance of preplate separation to the

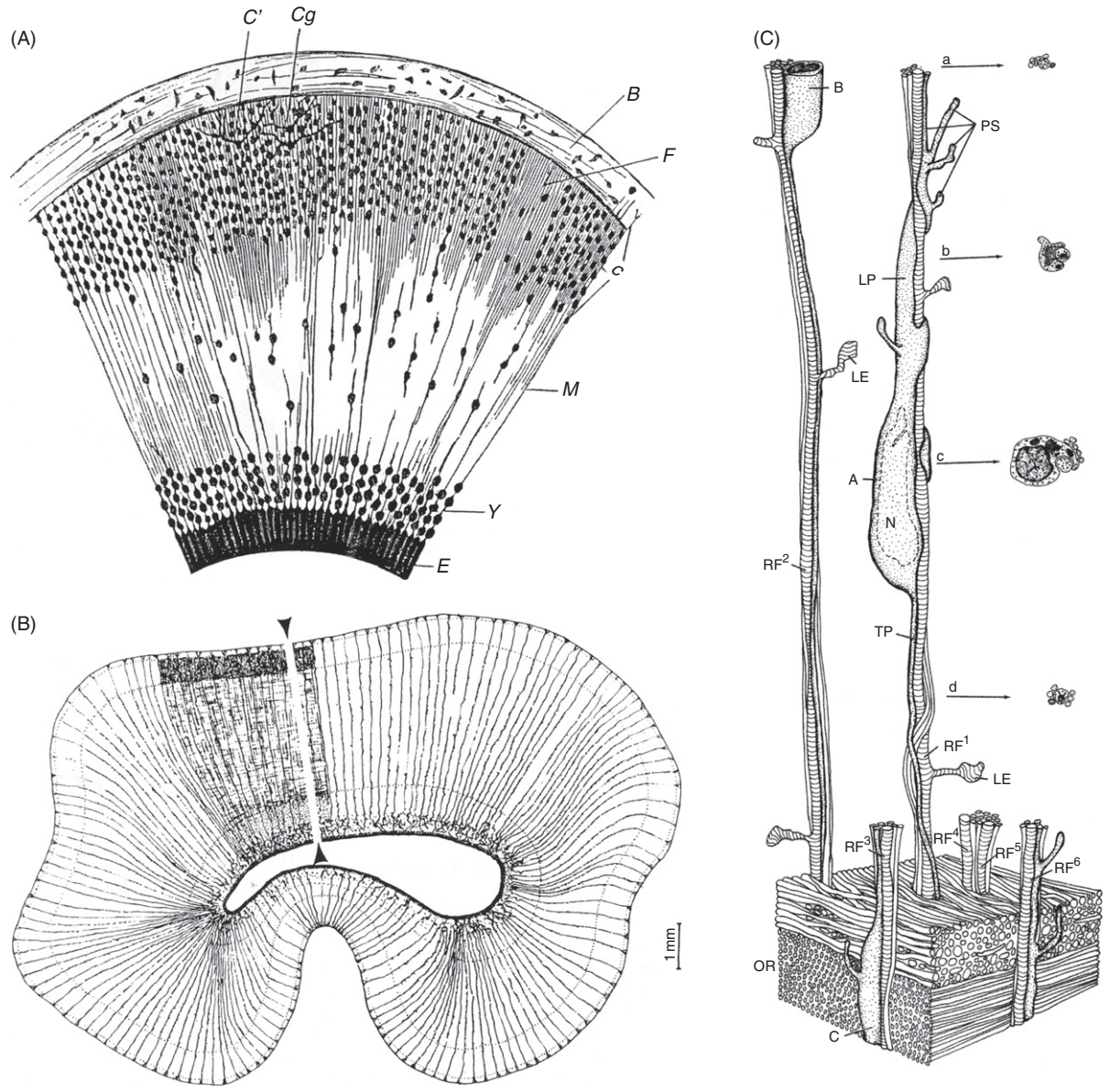


FIGURE 15.10 Radial Glial Cells Provide a Pathway for Neuronal Migration. (A) In the late nineteenth century, the Italian histologist G. Magini used Golgi impregnation methods to visualize a system of glial fibers that spanned the cortical wall. He proposed that the glial fibers provided a scaffold for neuronal migration. (B) In the late twentieth century, Rakic and Sidman again used the Golgi impregnation method to map radial glial cells in the developing primate neocortex. As described below, they used electron microscopy (EM) to visualize the relationship of young neurons to the glial fiber system. (C) Three-dimensional reconstruction of serial EM sections of a migrating neuron in the intermediate zone of the primate neocortex illustrate the cytology and neuron-glia relationships of migrating neurons *in vivo*. The cell soma of the migrating cell apposes radial glial fibers (striped vertical shafts, RF1-6), which extend short lamellate expansions (LE). Nuclei (N) of migrating neurons are elongated, and their leading processes (LP) are thicker and richer in organelles than their trailing processes (TP). The lower part of the diagram depicts the numerous parallel axons of the optic radiations (OR). These axons have been deleted from the upper portion of the figure to reveal the radial glial fibers. The leading process extends several pseudopodial endings (PS), which appear to explore the territory through which the neuron is migrating. In cross-section (a–d), a migrating neuron partially encircles the shaft of the radial glial fiber. *From Sidman and Rakic (1973).*

development of cortical laminae. During corticogenesis in *reeler* mutant mice, the segregation of preplate neurons into two zones and formation of the cortical plate do not occur and subsequent cortical lamination fails (Caviness & Rakic, 1978; Goffinet, 1984; Tissir & Goffinet, 2003). Cell movements involved in the subdivision of the preplate into the MZ and SP and formation of the CP are also critical to the formation of cortical synaptic circuits. As first proposed by Shatz and colleagues, subplate neurons extend pioneer axons during the establishment of thalamocortical architecture and function (Dupont, Hanganu, Kilb, Hirsch, & Luhmann, 2006; Ghosh & Shatz, 1992, 1993; Ghosh,

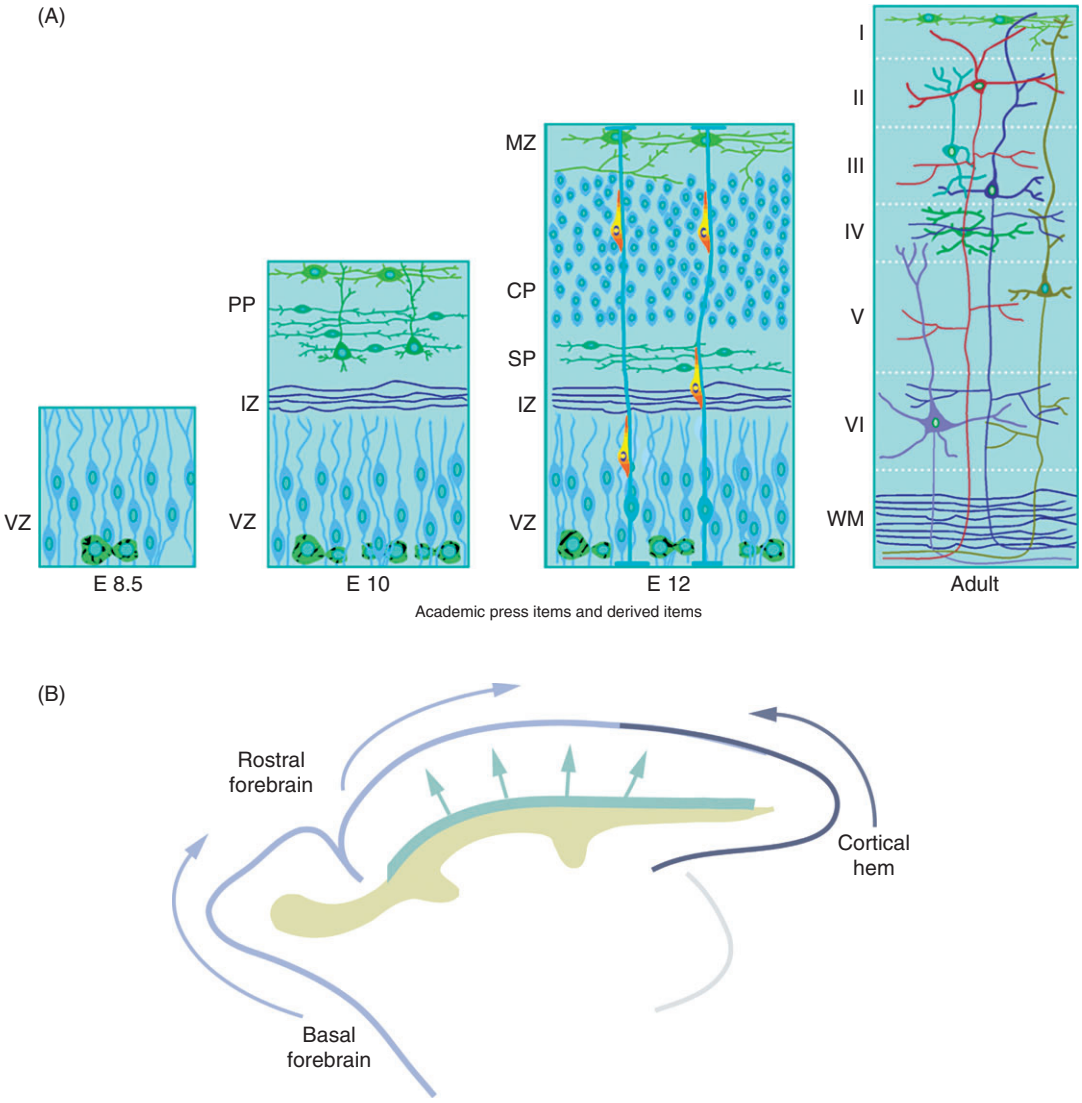


FIGURE 15.11 Development of the Cerebral Cortex. (A) *Formation of the Fundamental Layers*: The ventricular zone (VZ) contains the progenitors of neurons and glia. The first neurons to be generated establish the preplate (PP); their axons, as well as ingrowing axons from the thalamus, establish the intermediate zone (IZ). The subsequently generated neurons of cortical layers II–VI establish the cortical plate (CP), which splits the preplate into the marginal zone (MZ), or future layer I, and the subplate (SP), a transient population of neurons. After the completion of neuronal migration and differentiation, six cortical layers are visible overlying the white matter (WM) and the subplate has largely disappeared. Neural precursors in the subventricular zone (SVZ) continue to generate neurons that migrate rostrally into the olfactory bulb, even during postnatal life. (B) *Emerging Complexity of the Origin of Layer 1 Neurons*. Recent experiments in rodents and primates reveal multiple origins and migratory pathways of the precursor neurons in layer 1 of the developing neocortex. As discussed above, the neocortical VZ is a primary source of preplate neurons, which contains layer 1 cells. However, subsets of layer 1 neurons also arise in the rostral aspect of the neocortex, the basal forebrain, and the cortical hem. The latter progenitor cells migrate along the surface of the emerging brain from their sites of origin into the superficial layer of the neocortex. (This population of cells migrates several days before progenitors destined to become cortical interneurons). See text for details.

Antonini, McConnell, & Shatz, 1990; Kanold, Kara, Reid, & Shatz, 2003; Lein, Finney, McQuillen, & Shatz, 1999; Price et al., 2006). They form some of the first synapses in cortex (Blue & Parnavelas, 1983a, 1983b; Chun & Shatz, 1988; Konig & Marty, 1981; Konig, Roch, & Marty, 1975; Kostovic & Rakic, 1980; Molliver, Kostovic, & van der Loos, 1973) and provide a transient scaffold for the organization of thalamocortical projections. Thus, the neuroepithelium develops from a single layer of dividing cells to five embryonic zones, a VZ of proliferating progenitors, an intermediate zone (IZ) of immature axons, an SP containing a subset of preplate neurons, the cortical plate containing immature pyramidal cells born after embryonic day 13 (E13), and a superficial marginal zone (MZ), which contains Cajal-Retzius cells and small interneurons (Fig. 15.11).

Cortical Neurogenesis

Detailed cell cycle analyses by Takahashi and Caviness (Caviness et al., 2003) demonstrate that the murine neocortex is generated over an epoch of 11 cell cycles. Time-lapse imaging studies in cortical slices by McConnell and colleagues first suggested that progenitor cells can divide asymmetrically during the period of neurogenesis and that cleavage orientation of dividing precursors predicts the fate of the daughters (Chenn & McConnell, 1995). These studies proposed that vertically oriented divisions are symmetric, generating two identical daughters, whereas horizonal divisions are asymmetric, generating one daughter that differentiates into a neuron and another that reenters the cell cycle. The asymmetric inheritance of Notch and Numb proteins may also contribute to the control of symmetric versus asymmetric cell divisions during cortical neurogenesis. The polarity protein Numb is expressed in early phases of cortical histogenesis (Petersen, Zou, Hwang, Jan, & Zhong, 2002), and colocalizes with EGFR by a process that is actin-dependent (Sun, Goderie, & Temple, 2005). The loss of *Numb* and *Numbl* causes premature progenitor cell depletion and malformations of the neocortex (Petersen, Zou, Krauss, & Zhong, 2004), suggesting that *Numb*-mediated asymmetric cell divisions provide a general mechanism for cell cycle control during cell fate allocation events in the developing mammalian brain (Petersen et al., 2004).

At the cortical plate stage, two classes of neural stem and progenitor cells, the apical and basal progenitors (Farkas & Huttner, 2008; Gotz & Huttner, 2005; Noctor et al., 2002; Pinto & Gotz, 2007) are defined by their position in the cortical wall. The apical progenitors located in the VZ include the neuroepithelial cells (NECs) (Fig. 15.9) and radial glial cells (RGCs). RGCs are a specialized form of glial cell, recognized by their expression of the RC2 or BLBP antigen, that extend long radial processes toward the overlying cerebral wall (Gadisseux & Evrard, 1985; Kriegstein & Gotz, 2003; Noctor et al., 2002). As discussed below, these radial processes presage the basic columnar plan of development, providing a scaffold "in plane" with the layer of dividing cells for a subpopulation of neurons to migrate away from the primary germinal matrix. RGCs undergo symmetric divisions at early stages of cortical development, then shift to asymmetric divisions later, generating an RGC and either a neuron or a basal progenitor cell (also called an intermediate progenitor cell; IPC) that in turn divides in a more basal location, the subventricular zone (SVZ), and in most cases generates two neurons (Box 15.1, Fig. 15.14) (Farkas & Huttner, 2008; Gotz & Huttner, 2005; Kriegstein & Parnavelas, 2006; Zhong & Chia, 2008).

Cell Movements within the Primary Proliferative Matrix

The columnar organization of the cerebral cortex led to the hypothesis that progenitors in the neuroepithelium form a protomap for cortical organization. Two experimental approaches provided evidence for widespread tangential movements within the neuroepithelium. First, live imaging of dye-labeled cells in slices of embryonic cortex revealed extensive movements of the neuroblasts within the neuroepithelium. Labeled cells move intermittently within the plane of the epithelium at speeds between 10 and 100 μm/h (Fishell, Mason, & Hatten, 1993). Second, experiments using retroviral markers to follow the dispersion of clonally related cells in the developing brain demonstrated widespread dispersion of cells within the developing cortex (Walsh & Cepko, 1992). Thus, widespread movements occur in the neuroepthelial germinal zone of the CNS and across the plane of the outer embryonic zones of the neocortex.

Molecular Mechanisms of Cell Motility: General Features

Neuronal motility in the developing CNS, like that of all metazoan cells, occurs in response to a migration-promoting signal(s), which polarizes the cell and induces the extension of protrusions in the direction of migration (Ridley et al., 2003). These protrusions can be large, broad lamellipodia or spike-like filopodia driven by actin polymerization and microtubule dynamics, and stabilized by adhering to the extracellular matrix (ECM) in the case of migrating neural crest cells or adjacent to radial glial cells or to bundles of axons in the case of radial, glial-guided migrations and tangential migrations. Adhesions serve as traction sites during forward translocation of the cell soma (Fig. 15.13).

As key regulators of the actin and microtubule cytoskeletons, the Rho family of small GTPases plays an integral part in the morphological changes, polarization, and locomotion of migrating cells. When bound to GTP, they stimulate downstream effector molecules to regulate actin and microtubule dynamics to reorganize the cytoskeleton, as well as control cell-cycle progression, gene expression, and vesicle transport. The best-studied Rho GTPase family members are Rac, Cdc42, and RhoA. In many cell types, Rac stimulates the formation of the veil-like lamellipodial protrusions involved in membrane extension, Cdc42 promotes the formation of the thin filopodial microspikes that sense the surrounding environment, and RhoA regulates cell contractility. All three of these GTPases play a role in cell adhesion (Ridley et al., 2003). Cdc42 also has a conserved role in regulating cell polarity, along with homologs of the *C. elegans* polarity signaling complex Par6, and together they orient the nucleus, centrosome, and Golgi apparatus in the direction of migration (Burakov, Nadezhdina, Slepchenko, & Rodionov, 2003; Etienne-Manneville & Hall, 2001; Etienne-Manneville & Hall, 2003; Gomes, Jani, & Gundersen, 2005). The Rho GTPases orchestrate general and specific structural and functional

BOX 15.1

ADULT NEUROGENESIS

Adult neurogenesis refers to the process of generating new neurons in the adult mammalian brain (Gage, 2000). For over a century, it was believed that adult mammalian brains had no capacity to produce new neurons and that the neurons damaged in the adult brain were lost forever. However, this dogma has been challenged by the discovery of neural stem cells (NSCs) in the adult brain and by intensive research on adult neurogenesis over the past two decades. These studies of adult neurogenesis can be dated back to the 1960s, when Joseph Altman and colleagues observed [^{3}H]-labeled neurons in brains of adult rodents injected with [^{3}H]-thymidine, a marker for proliferative cells and their progeny (Altman & Das, 1965). A second breakthrough in the study of adult neurogenesis did not occur until the early 1990s, however, with the discovery of adult NSCs that are able to self-renew while producing daughter cells with a potential to differentiate into neurons (Reynolds & Weiss, 1992). Two discrete neurogenic niches that harbor NSCs have been consistently identified across mammalian species in vivo—the subventricular zone (SVZ) of the lateral ventricle and the subgranular zone (SGZ) in the dentate gyrus of the hippocampus. NSCs in the SVZ give rise to neuronal precursors that migrate through the rostral migratory stream and differentiate into interneurons in the granular cell layer and the periglomerular layer in the olfactory bulb. In the SGZ, most of the progeny of NSCs differentiate and integrate locally as dentate granule cells. Adult neurogenesis in the hippocampus can also be observed in human brains (Eriksson et al., 1998), suggesting a therapeutic potential for utilizing adult-born neurons for treatment of neurodegenerative disease and repair of injured neural tissues in adulthood.

Adult neurogenesis starts with the proliferation of NSCs. Some progeny of NSCs remain undifferentiated as NSCs; others differentiate into neurons and glia. The newly born neurons undergo a lengthy process of morphological and physiological maturation and eventually integrate into the corresponding neural circuitry. During the differentiation and maturation processes, a significant fraction of newborn neurons are lost before their final integration. Therefore, the rate of neurogenesis is regulated at levels of NSC proliferation, neuronal fate determination, and the survival of newborn neurons. Numerous molecular factors take part in the regulation of adult neurogenesis, including growth factors, neurotrophins, morphogens, cytokines, and hormones. Moreover, distinct from embryonic neurogenesis, adult neurogenesis can also be modulated by the animals' experiences and behaviors. For example, voluntary exercise, environmental enrichment, and hippocampus-dependent learning can enhance SGZ neurogenesis. Conversely, hippocampal neurogenesis is reduced in stressed or aged animals. Olfactory stimulus enrichment results in enhanced SVZ neurogenesis.

Because of the region-specific network integration of newborn neurons, it is postulated that SGZ neurogenesis may play a role in hippocampus-related functions, such as learning and memory and affective regulation, and SVZ neurogenesis may contribute to olfactory function. This hypothesis is supported by many studies demonstrating a positive correlation between levels of neurogenesis and performance in specific behavioral tasks. While a causal relationship between adult neurogenesis and its functions is the subject of intense investigation, current studies using a combination of molecular, cellular, genetic, system, and behavioral approaches are striving to elucidate the exact roles of newborn neurons in hippocampal and olfactory functions.

Wei Deng and Fred H. Gage

changes required for directed motility of nonneuronal cells and neurons, as well as growth cones.

A large number of actin regulatory molecules act downstream of Rac and Cdc42 signaling to regulate the formation of lamellipodia and filopodia, respectively, during motility. Rac induces actin filament uncapping, actin polymerization, and filament branching during lamellipodial extension, and Cdc42 induces the actin polymerization of parallel bundles of actin filaments that make up filopodia. Cdc42 and Rac localize Wiskott–Aldrich syndrome protein (WASP) family members and WASP-family verprolin-homologous protein (WAVE) complexes, respectively, at the cell surface to activate the actin-related protein-2/3 (ARP2/3) complex. Arp2/3 is an actin-nucleating complex that binds to the sides or tip of a preexisting actin filament and induces the formation of a new daughter filament that branches off the mother filament (Heasman & Ridley, 2008; Ridley et al. 2003). Protrusive force by actin polymerization is generated by an "elastic Brownian ratchet" mechanism, in which thermal energy bends nascent short actin filaments, pushing the membrane forward. Thus, actin polymerization and unbending a growing actin filament against the leading edge of the cell provides a driving force for the motility of nonneuronal cells, neural crest cells and extending axons. Local activation of the Arp2/3 complex induces lamellipodial extension in a particular direction, providing the basis for directional migration.

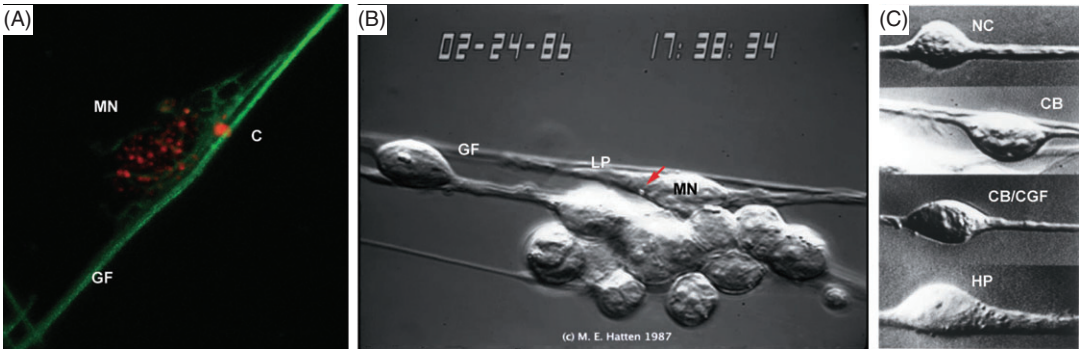


FIGURE 15.12 Live Imaging of Glial-Guided Neuronal Migration. Live imaging of cells with video-enhanced differential contrast (VEC-DIC) optics. The cytology and neuron-glial relationship of a cerebellar granule neuron migrating on a glial fiber. With live imaging of neuronal migration along glial fibers, the granule neuron closely apposes the cell glial fiber (GF) along the length of the soma of the migrating neuron (MN) and extends a leading process (LP) in the direction of migration along the glial guide. The nucleus is in the posterior aspect of the cell during periods of movement, the centrosome (C) is just forward of the neuronal nucleus. (A) In migrating neurons, apparently antibodies against dynein intermediate chain (red) label the centrosome and dynein on the surface of the nucleus. Antibodies against tubulin (green) reveal a perinuclear "cage" of tubulin in migrating neurons. (B) Live imaging of a cerebellar granule neuron migrating on a glial fiber (red arrow indicates centrosome). (C) Cytology of cerebellar and hippocampal neurons migrating along homotypic and heterotypic astroglial fibers: cerebellar granule neuron migrating along a hippocampal glial fiber (CB/HP), a cerebellar granule neuron migrating on a cerebellar glial fiber (CB/CB), a hippocampal neuron migrating along a cerebellar glial fiber (HP/CB) or a hippocampal neuron migrating on a hippocampal glial fiber (HP/HP). In heterotypic co-cultures, migrating neurons have a stereotyped cytology, extending their cell soma along the glial guide and extending a leading process in the direction of migration.

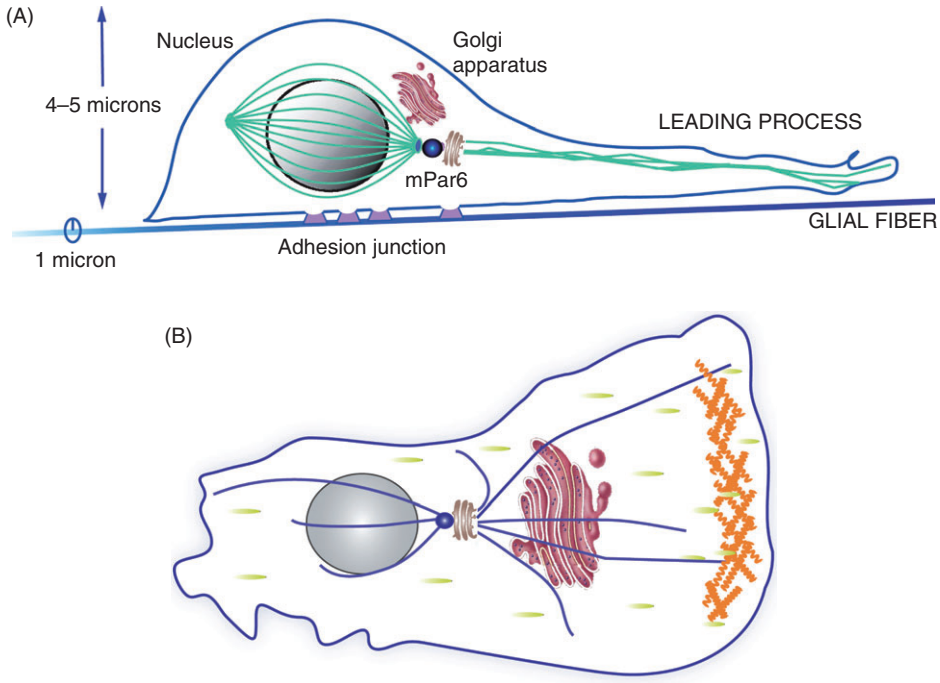


FIGURE 15.13 *Comparison of the cellular organization of neurons migrating on glial fibers (A) and fibroblasts migrating in vivo (B).* As the neuron migrates, the polarity complex mPar6, and signaling pathways, polarize the migrating cell. mPar6 localizes to the centrosome (blue) and paces the timing of the movement of the neuron along the glial fiber via Myosin IIB-mediated contraction of F-actin in the proximal aspect of the leading process. A perinuclear cage of tubulin (green) positions the nucleus in the posterior aspect of the cell. Microtubules (green) grow out of the centrosome into the leading process and support vesicle transport during movement. As the neuron moves, it forms attachments beneath the cell soma (purple), which provide traction for neuronal locomotion along the fiber. (B) The model shown represents a polarized cell that has distinct leading and trailing edges. This is a common feature of fibroblastic motility. The leading edge points in the direction of movement and is driven by actin-polymerization-mediated protrusion. Green spots represent points of interaction of the cell with the substrate. Other structures depicted include the nucleus (grey), the Golgi apparatus (brown) and the microtubule-organizing center (MTOC) (blue), from which the microtubule network (blue) radiates, as well as an actin-rich lamellipodium at the front (orange).

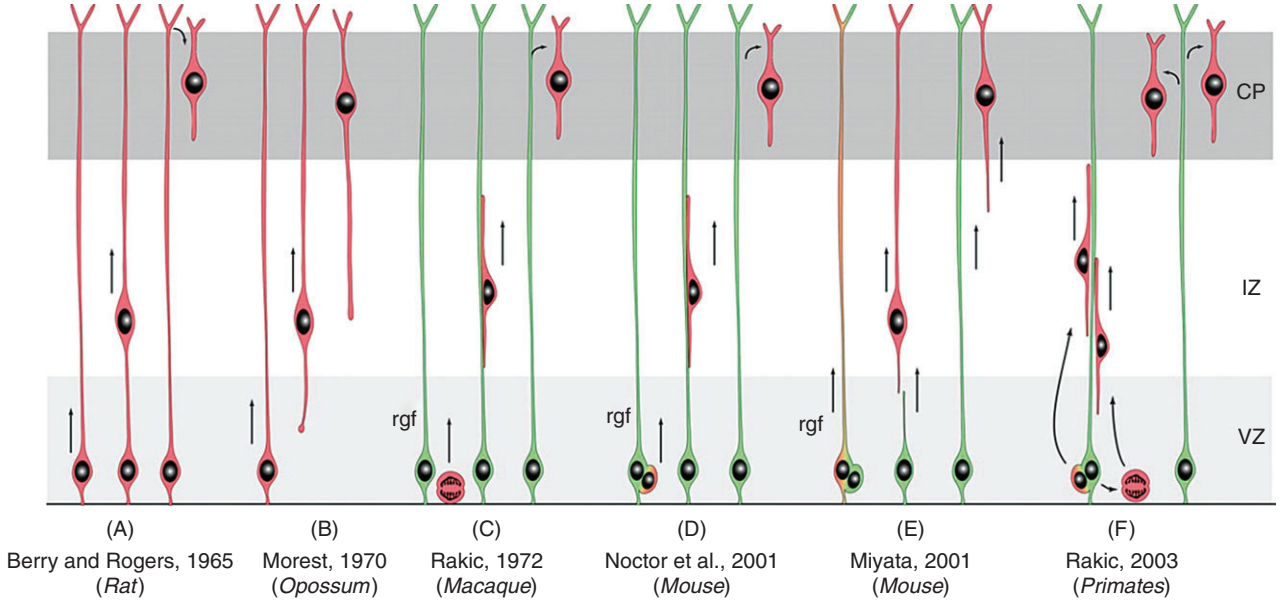


FIGURE 15.14 Schema of the Evolving Concepts of the Relationship between Radial Glial Cells and Migrating Neurons in the Developing Mammalian Cerebral Cortex. As discussed in the text, initial theories of neuronal migration ignored the glial cells. In the early 1970s, Sidman and Rakic proposed that neurons migrate along glial fibers. Live imaging of labeled radial glia and migrating neurons revealed that during the epoch of migration, glial cells produce neuronal progeny, which immediately or after several rounds of division use the mother glial cell as a guide for their migration. As indicated in the diagram, there are some differences in this process among vertebrate species. Neurons, red; radial glial cells, green. *Reprinted from Rakic (2003).*

Neuronal Migration along Radial Glial Fibers

Radial glial cells are a transient population of cells that provide a primary pathway for neuronal migrations. The histologists Albert Kölliker and Wilhelm His used silver impregnation methods to reveal the development of glial cells in the late nineteenth century. In a series of EM and Golgi studies, Rakic and Sidman provided evidence that neurons are closely apposed to radial glial fibers during the period of neuronal migration (Rakic, 1972; Sidman & Rakic, 1973) (Fig. 15.10). Rakic and Sidman postulated that the radial glia guide the migrations of young neurons from the VZ to the neural layers, in a spatiotemporal order that positioned the youngest born neurons in the most superficial layers. This "inside-out" model relied upon coordinating ^{3}H-thymidine labeling of neuroblast proliferation in the VZ with light and EM analysis (Sidman, 1970). Cells born at successively later times after administration of labeled ^{3}H-thymidine would be lighter than those born the day of the injection (their "birthday"), due to dilution of the label with continued proliferation.

Experiments by Nowakowski and Rakic (1979) on the developing hippocampus provided the most compelling argument for glial guidance of migration, as young neurons in the hippocampal formation followed the undulating paths of the glial fibers, rather than a straight radial line, during their migrations. Real-time imaging of the migration of identified neurons, purified from cerebellum, hippocampus, and neocortex, on glial processes provided critical support for this model (Edmondson & Hatten, 1987), and revealed the dynamics of cell movement along the glial guide described above. Cell-based assays further showed that neurons from one brain region could migrate freely on glia from other brain regions, suggesting that the glial fibers were a generic "monorail" system, rather than a track that specified the duration and direction of migration (Gasser & Hatten, 1990a, 1990b; Hatten, 1999) (Fig. 15.12).

Live imaging studies of neuronal migration along glial fibers revealed unique features of migrating neurons that are distinct from features of general cell and axon motility. These unique features include the extension of a highly polarized leading process in the direction of migration, the assembly of an interstitial adhesion junction beneath the cell soma (Gregory, Edmonson, Hatten, & Mason, 1988), localization of acto-myosin contractile motors in the proximal portion of the leading process (Solecki et al., 2009), and the formation of a perinuclear cage of tubulin to maintain the posterior positioning of the nucleus (Fig. 15.13). Migration involves a cycle of steps that include forward movement of the centrosome into the proximal portion of the leading process, translocation of the neuronal nucleus (Edmondson & Hatten, 1987; Solecki, Model, Gaetz, Kapoor, & Hatten, 2004), activation of acto-myosin motors in the proximal aspect of the leading process (Solecki et al., 2009), and the release of the adhesion junction. Thereafter, the neuronal cell soma glides forward along the glial fiber until a new adhesion forms.

During glial-guided neuronal migration, Cdc42 likely polarizes migrating neurons, while Rac regulates leading process formation and protrusion. A decrease in RhoA levels and activity are required for radial migration; however, low levels likely persist to regulate actomyosin contractility to generate force during motility (Govek, Hatten, & Van Aelst, 2011). Recent studies show that two additional Rho GTPase proteins, Rnd2 and Rnd3/RhoE, inhibit RhoA signaling during radial migration to control distinct steps of the migratory process (Heng et al., 2008; Pacary et al., 2011). Thus, tight regulation of Rho GTPase levels and activity are required for proper neuronal migration, and the Rho GTPases act within multiple signaling pathways, including those associated with cortical malformations, to regulate different steps of the migration cycle.

During migration, a dynamic rearrangement of the actin cytoskeleton occurs in the proximal region of the leading process, where Myosin II motors localize. Myosin II activity, which is regulated by the conserved mPar6α polarity complex, is required for both centrosomal and somal motility (Solecki et al., 2004, 2009). The microtubule cytoskeleton is further regulated by cytoplasmic dynein and its cofactors LIS1 and Doublecortin (DCX) to couple centrosomal and nuclear movement (Coquelle et al., 2002; Faulkner et al., 2000; Mesngon et al., 2006; Smith et al., 2000; Tai, Dujardin, Faulkner, & Vallee, 2002; Tsai, Bremner, & Vallee, 2007; Tsai, Chen, Kriegstein, & Vallee, 2005; Umeshima, Hirano, & Kengaku, 2007).

Several adhesion receptor systems function in glial-guided neuronal migration, including Astrotactin (ASTN1), Neuregulin/ErbB4, and BDNF. The neuronal protein ASTN1 is a well-studied adhesion receptor for neuronal migration along glial fibers (Edmondson, Liem, Kuster, & Hatten, 1988; Fishell & Hatten, 1991) that is expressed by neurons in both the cerebellum and cerebral cortex (Zheng, Heintz, & Hatten, 1996). Mice lacking ASTN1 have slowed cerebellar granule neuron migration, a decrease in neuron-glial binding, and abnormal Purkinje cell development compared with wild-type littermates (Adams, Tomoda, Cooper, Dietz, & Hatten, 2002). Recent studies have identified a second member of the Astrotactin receptor family, ASTN2, which regulates the cell surface expression and trafficking of ASTN1 critical for granule cell migration (Wilson, Fryer, & Hatten, 2010). The importance of receptor trafficking to neuronal migration is further highlighted by the finding that inhibition of endocytosis by the small molecule dynasore reversibly arrests migration (Wilson et al., 2010; Zhou et al., 2007). Neuregulin binds ErbB4 on the glial surface and provides signals that maintain glial process formation (Anton, Marchionni, Lee, & Rakic, 1997; Rio, Rieff, Pelmin, & Corfas, 1997), and BDNF stimulates granule neuron migration (Borghesani et al., 2002). Surprisingly integrin-based adhesions, which are essential for most types of cell migration, are not required for glial-guided neuronal migration in the developing brain (Belvindrah, Hankel, Walker, Patton, & Muller, 2007; Fishell & Hatten, 1991).

As methods became available to culture acute slices of developing brain, it became feasible to acquire real-time images of neurons migrating in the complex setting of developing brain tissue. Lipophilic dyes (Rivas & Hatten, 1995), retroviral transfection, and *in vivo* electroporation of constructs that tag the cells with fluorescent markers such as EGFP (Bhatt, Tomoda, Fang, & Hatten, 2000) in the embryonic neocortex, followed by the generation of acute slices several days later, made it possible to follow particular cohorts of developing neurons and their fates in the emerging neocortex (Zhang, Campbell, Ting, & Tsien, 2002). Moreover, this approach facilitated real-time imaging of cells with altered levels of gene expression, achieved by the introduction of cDNAs to overexpress genes or small hairpin RNAs (shRNAs) to knockdown gene expression, in order to test the function of proteins associated with human cortical malformations (Bai et al, 2003; Tsai et al., 2005), transcription factors (Hand et al., 2005), polarity proteins (Solecki et al., 2004), cytoskeletal proteins (Tsai et al., 2005), and components of downstream signaling pathways. Methods for real-time imaging, such as spinning disc confocal imaging, and for imaging cells deep in living tissue with multiphoton systems are evolving rapidly, providing neurobiologists with the tools to study the molecular regulation of CNS migrations. These experiments have occurred in parallel with cell biological studies of the fundamental mechanisms of metazoan motility, described briefly above.

Lineage Analysis of Radial Glial Cells: Glia Generate Neurons in Late Phases of Cortical Development

During the past decade, real-time imaging of labeled radial glial cells and genetic experiments on the lineal descent of radial glia have shown that a large number of neurons are descended from glial progenitor cells. In the developing neocortex, imaging experiments (Noctor et al., 2002; Kriegstein & Noctor, 2004) revealed the surprising finding that radial glial cells can also generate neuronal precursors, which immediately, or after several rounds of cell division, migrate along the radial processes of the glial mother cells that stretch across the expanding cerebral wall (Fig. 15.14). Genetic studies by Anthony and Heintz (Anthony, Klein, Fishell, & Heintz, 2004) used Cre/loxP fate mapping and clonal analysis to demonstrate that radial glia throughout the CNS serve as neuronal progenitors and that radial glia within different regions of the CNS pass through their neurogenic stage of development at distinct time points. Thus, radial glial populations within different CNS regions are not heterogeneous with regard to their potential to generate neurons versus glia. They went on to show that the promoter region of the radial glial cell marker gene *Blbp* (Feng, Hatten, & Heintz, 1994) contains a binding site for the Notch effector CBF1

that is essential for *Blbp* transcription in radial glia (Anthony, Mason, Gridley, Fishell, & Heintz, 2005). These results identified *Blbp* as the first predominantly CNS-specific Notch target gene. Their results also confirmed earlier studies by Gaiano and Fishell (Gaiano & Fishell, 2002) who showed that Notch signaling regulates radial glial differentiation in the neocortex.

Tangential Migration Pathways: Cortical Interneurons, Cajal-Retzius Cells, the Rostral Migratory Stream, and the Cerebellar System

CORTICAL INTERNEURONS

Although radial migration was once thought to be the primary mode of CNS migrations in cortical regions of vertebrate brain, we now realize that a small subset of the total population of neurons in the murine neocortex migrate from the neocortical VZ to form the layers of the mature cerebral cortex (Tan et al., 1995, 1998). These neurons are the large output neurons, the pyramidal cells (Parnavelas, Barfield, Franke, & Luskin, 1991), and a subpopulation of cortical interneurons. The majority of cortical interneurons originate in the basal forebrain and migrate into the neocortex (Fig. 15.15). The importance of this ventral to dorsal mode of migration is evidenced by the rather remarkable fact that the GABAergic neurons are only about 20% of the total neuronal population in the cortex of rodents (the number is far lower in humans). A detailed view of the migratory routes of cells migrating from the basal forebrain into the neocortex emerged from genetic and transplantation studies. Initial experiments with mice lacking the transcription factors *Dlx1/2* suggested that cortical interneurons originate in the medial ganglionic eminence (MGE) (Anderson, Eisenstat, Shi, & Rubenstein, 1997; Tamaki, Fujimori, & Takauji, 1997). More detailed transplantation and fate mapping studies refined this original view, providing support for the general model that the majority of cortical interneurons originate in the MGE and migrate into the neocortex (Polleux, Whitford, Dijkhuizen, Vitalis, & Ghosh, 2002; Xu, Cobos, De La Cruz, & Rubenstein, & Anderson, 2004). The lateral ganglionic eminence (LGE) likely contributes a small population of interneurons to the cortical interneuron population, but current experiments show that the many LGE cells migrate into the olfactory bulb intermingling with cells that migrate from the cortical SVZ.

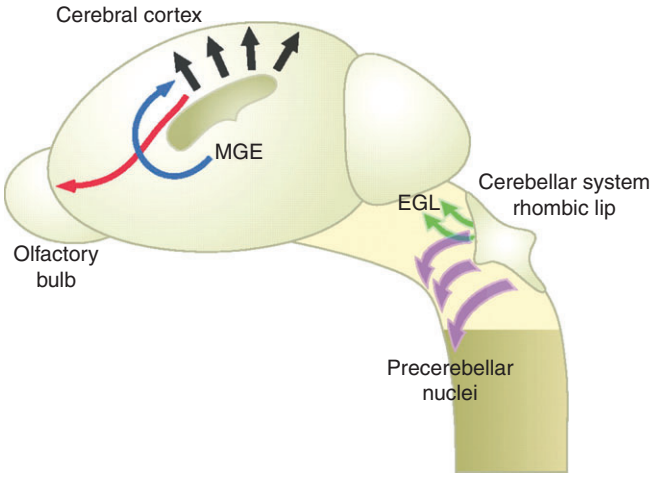


FIGURE 15.15 Non-Radial Pathways of Neuronal Migration. Over the past decade, a number of tangential migratory pathways have been described, in addition to radial pathways (black), including the migration of interneurons from the basal forebrain into the developing neocortex (blue), the migration of neurons from the SVZ to the olfactory bulb along the rostral migratory stream (red), and migrations of progenitors from the rhombic lip to cerebellar cortex (green) and the pre-cerebellar nuclei of the brainstem (purple). (EGL, external granule layer; MGE, medial ganglionic eminence; SVZ, subventricular zone) *Reprinted from Hatten (2002).*

Another tangential migratory pathway occurs in the marginal zone of the neocortex. Although the cells of the MZ were long assumed to be the earliest generated neurons in the cortical VZ, recent evidence from genetic and molecular genetic studies supports the view that a subpopulation of MZ cells originate in the rostral forebrain (or even the basal forebrain) (Meyer, Soria, Martinez-Galan, Martin-Clemente, & Fairen, 1998), while others emerge from the cortical hem, and migrate across the surface of the developing cortex (Garcia-Moreno, Lopez-Mascaraque, & De Carlos, 2007; Meyer, 2007). Support for this view came from studies on gene expression patterns in EGFP-BAC transgenic mice from the GENSAT project (Gong et al., 2003), which showed that cells marked by the gene *Pde1c* migrate from the rostral forebrain back over the surface of the emerging neocortex. Similarly, studies on *Wnt3a positive* cells showed that cells of the cortical hem also contribute to the preplate via a superficial migratory pathway across the surface of the neocortex (Louvi, Yoshida, & Grove, 2007).

THE ROSTRAL MIGRATORY STREAM

The rostral migratory stream (RMS) is another major pathway for tangential migrations (Fig. 15.5). Between E12.5 and E14 in the mouse (E14-16 in the rat), a secondary proliferative zone is formed along the third ventricle of the forebrain (Pencea & Luskin, 2003). This progenitor zone, known as the subventricular zone (SVZ), persists in the adult, where it continues to generate neurons. The subventricular zone has been the subject of intense interest, because it continues to generate new neurons into adulthood, providing GABAergic neurons for the olfactory bulb (Luskin, 1998). As a renewable population of partially committed CNS precursors, the cells from the RMS are an ideal system for studying adult neural "stem" cells (as indicated, they are not "stem cells" in the strict definition of the term). RMS cells differentiate into interneurons when transplanted into the wide variety of target areas, including the septum, thalamus, hypothalamus, and

midbrain of embryonic brain (Alvarez-Buylla, Herrera, & Wichterle, 2000). Thus, many labs are studying the molecular mechanisms that control the differentiation of these cells when placed in ectopic positions, especially in the adult brain.

RMS neurons migrate from the SVZ to the olfactory bulb by a unique mechanism. As the neurons move, they use other migrating neurons in the stream as a substrate (Lois & Alvarez-Buylla, 1994; Lois, Garcia-Verdugo, & Alvarez-Buylla, 1996). This mode of neuronal migration is termed "chain migration" because the neurons form chains within glial sheaths. Recent experiments by Muller and colleagues show that neuroblasts express integrins containing the ß1 and ß5 subunits (Belvindrah et al., 2007). Using genetic and cell biological approaches, they demonstrated that ß5 integrins are dispensable for chain migration, whereas ß1 integrins promote cell–cell interactions that link neuroblasts into chains. The ß1 integrin ligand laminin is recruited to the cell surface of migrating neuroblasts and induces chain formation in SVZ explants and the aggregation of purified neuroblasts. Thus, ß1 integrins and their laminin ligands promote the formation of cell chains in the adult RMS (Belvindrah et al., 2007).

Several secreted signaling molecules are thought to guide the migration of RMS neuroblasts to the olfactory bulb, including the diffusible chemoattractants Netrin-1 (Murase & Horwitz, 2002), the glial cell line–derived neurotrophic factor (Paratcha, Ibanez, & Ledda, 2006), and the chemorepellent Slit1/2 (Nguyen-Ba-Charvet et al., 2004; Wu et al., 1999). In addition, receptor tyrosine kinases of the Eph families and their ephrin ligands (Conover et al., 2000), and Neuregulin and Erb-B ligands (Ghashghaei et al., 2006) apparently guide neuronal migration in the RMS.

CEREBELLAR SYSTEM

A final example of widespread tangential migrations can be found in the cerebellar system, where migrations from the rhombic lip (RL) contribute to both cerebellar histogenesis and the formation of the precerebellar nuclei of the brainstem, which project afferent connections to the cerebellar nuclei or receive efferent connections from the cerebellum (Fig. 15.5). During embryogenesis, the cerebellar territory arises from rhombomere 1, delineated by the expression of Hoxa2 (posterior) and Otx2 (anterior). In the mouse, between embryonic (E) day E9.0 and E9.5, the pontine flexure creates an area along the border of the mesencephalon and metencephalon, where the neural tube fails to close, called the RL (Harkmark, 1954). The embryonic RL is a specialized germinative epithelium that arises relatively late in development at the interface between the neural tube and the roofplate of the fourth ventricle. The rostral (cerebellar) and hindbrain domains of the RL form a continuous proliferative epithelium, which contains overlapping pools of progenitor cells that express combinatorial patterns of transcription factors. Lineage tracing experiments demonstrate that progenitor cell populations in the RL generate both cerebellar granule cells and neurons of the cerebellar nuclei (Rodriguez & Dymecki, 2000), the lateral pontine nucleus, and the cochlear nucleus (Machold & Fishell, 2005; Wang, Rose, & Zoghbi, 2005). Moreover, lineage tracing experiments define the neurons of the precerebellar nuclei that project afferent axons to cerebellar neurons (Rio et al., 1997). Thus, experiments on the allocation of cell fate among rhombic lip (RL) precursors destined for the cerebellar cortex, the cerebellar nuclei and the pre-cerebellar nuclei of the chick and mouse (Landsberg et al., 2005) illustrate the complex migrations that generate the nuclear and cortical regions of the cerebellum (Morales & Hatten, 2006). These patterns include both radial and tangential migratory pathways.

Summary

During CNS development, a remarkable range of neuronal migrations set forth the architectonics of the brain. In general, these migrations can be seen as dorso-ventral (DV) or anterior-posterior (AP) migrations, pathways thought to be prominent in lower organisms but not in vertebrates. Indeed, genes discovered in *C. elegans* and *Drosophila* provide molecular cues for guiding migrations in higher vertebrates, and conserved polarity proteins and their signaling pathways play a key role in sensing directional cues, polarizing cellular extrusions, and generating forces required to propel cells along pathways defined by specific adhesions and/or chemoattractants. The relationship of cell migrations to the establishment of the neuronal layers and brain circuitry will ultimately require novel insights into the role of the transient scaffolding systems of cortical development, the preplate, subplate, and cortical plate, which appear to pattern the connections between laminar, cortical areas of the brain, subcortical nuclei, and the formation of sensorimotor pathways.

References

Adams, N. C., Tomoda, T., Cooper, M., Dietz, G., & Hatten, M. E. (2002). Mice that lack astrotactin have slowed neuronal migration. *Development, 129,* 965–972.

Altman, J., & Das, G. D. (1965). Autoradiographic and histological evidence of postnatal hippocampal neurogenesis in rats. *Journal of Comparative Neurology, 124,* 319–335.

Alvarez-Buylla, A., Herrera, D. G., & Wichterle, H. (2000). The subventricular zone: Source of neuronal precursors for brain repair. *Progress in Brain Research, 127,* 1–11.

Anderson, S. A., Eisenstat, D. D., Shi, L., & Rubenstein, J. L. (1997). Interneuron migration from basal forebrain to neocortex: Dependence on Dlx genes. *Science, 278,* 474–476.

Anthony, T. E., Klein, C., Fishell, G., & Heintz, N. (2004). Radial glia serve as neuronal progenitors in all regions of the central nervous system. *Neuron, 41,* 881–890.

Anthony, T. E., Mason, H. A., Gridley, T., Fishell, G., & Heintz, N. (2005). Brain lipid-binding protein is a direct target of Notch signaling in radial glial cells. *Genes & Development, 19,* 1028–1033.

Anton, E. S., Marchionni, M. A., Lee, K. F., & Rakic, P. (1997). Role of GGF/neuregulin signaling in interactions between migrating neurons and radial glia in the developing cerebral cortex. *Development, 124,* 3501–3510.

Bai, J., Ramos, R. L., Ackman, J. B., Thomas, A. M., Lee, R. V., & LoTurco, J. J. (2003). RNAi reveals doublecortin is required for radial migration in rat neocortex. *Nature Neuroscience, 6,* 1277–1283.

Barkovich, A. J., Kuzniecky, R., Dobyns, W. B., Jackson, G., Becker, L. E., & Evrard, P. (1996). Malformations of the cortical development. *Neuropediatrics, 27,* 59–63.

Basch, M., Bronner-Fraser, M., & Garcia-Castro, M. (2006). Specification of neural crest occurs during gastrulation and requires Pax7. *Nature, 441,* 218–222.

Belvindrah, R., Hankel, S., Walker, J., Patton, B. L., & Muller, U. (2007). Beta1 integrins control the formation of cell chains in the adult rostral migratory stream. *Journal of Neuroscience, 27,* 2704–2717.

Berry, M., & Rogers, A. W. (1965). The migration of neuroblasts in the developing cerebral cortex. *Journal of Anatomy, 99,* 691–709.

Bhatt, R. S., Tomoda, T., Fang, Y., & Hatten, M. E. (2000). Discoidin domain receptor 1 functions in axon extension of cerebellar granule neurons. *Genes & Development, 14,* 2216–2228.

Blue, M. E., & Parnavelas, J. G. (1983). The formation and maturation of synapses in the visual cortex of the rat. I. Qualitative analysis. *Journal of Neurocytology, 12,* 599–616.

Blue, M. E., & Parnavelas, J. G. (1983). The formation and maturation of synapses in the visual cortex of the rat. II. Qualitative analysis. *Journal of Neurocytology, 12,* 697–712.

Borghesani, P. R., Peyrin, J. M., Klein, R., Rubin, J., Carter, A. R., Schwartz, P. M., …, Segal, R. A. (2002). BDNF stimulates migration of cerebellar granule cells. *Development, 129,* 1435–1442.

Bronner-Fraser, M. (1986). Analysis of the early stages of trunk neural crest migration in avian embryos using the monoclonal antibody HNK-1. *Developmental Biology, 115,* 44–55.

Bronner-Fraser, M., & Fraser, S. (1988). Cell lineage analysis shows multipotentiality of some avian neural crest cells. *Nature, 335*(8), 161–164.

Burakov, A., Nadezhdina, E., Slepchenko, B., & Rodionov, V. (2003). Centrosome positioning in interphase cells. *Journal of Cell Biology, 162,* 963–969.

Caviness, V. S., Jr. (1982). Neocortical histogenesis in normal and reeler mice: A developmental study based upon [3H]thymidine autoradiography. *Brain Research, 256,* 293–302.

Caviness, V. S., Jr., & Rakic, P. (1978). Mechanisms of cortical development: A view from mutations in mice. *Annual Review of Neuroscience, 1,* 297–326.

Caviness, V. S., Jr., Goto, T., Tarui, T., Takahashi, T., Bhide, P. G., & Nowakowski, R. S. (2003). Cell output, cell cycle duration and neuronal specification: a model of integrated mechanisms of the neocortical proliferative process. *Cerebral Cortex, 13,* 592–598.

Chenn, A., & McConnell, S. K. (1995). Cleavage orientation and the asymmetric inheritance of Notch1 immunoreactivity in mammalian neurogenesis. *Cell, 82,* 631–641.

Christensen, J. H., Coles, E. G., & Wilkinson, D. G. (2000). Molecular control of neural crest formation, migration and differentiation. *Current Opinion in Cell Biology, 12,* 719–724.

Chun, J. J., & Shatz, C. J. (1988). Redistribution of synaptic vesicle antigens is correlated with the disappearance of a transient synaptic zone in the developing cerebral cortex. *Neuron, 1,* 297–310.

Conover, J. C., Doetsch, F., Garcia-Verdugo, J. M., Gale, N. W., Yancopoulos, G. D., & Alvarez-Buylla, A. (2000). Disruption of Eph/ephrin signaling affects migration and proliferation in the adult subventricular zone. *Nature Neuroscience, 3,* 1091–1097.

Coquelle, F. M., Caspi, M., Cordelieres, F. P., Dompierre, J. P., Dujardin, D. L., Koifman, C., …, Reiner, O. (2002). Lis1, CLIP-170's key to the dynein/dynactin pathway. *Molecular and Cellular Biology, 22,* 3089–3102.

Del Rio, J. A., Martinez, A., Auladell, C., & Soriano, E. (2000). Developmental history of the subplate and developing white matter in the murine neocortex. Neuronal organization and relationship with the main afferent systems at embryonic and perinatal stages. *Cerebral Cortex, 10,* 784–801.

Dobyns, W. B., & Truwit, C. L. (1995). Lissencephaly and other malformations of cortical development: 1995 update. *Neuropediatrics, 26,* 132–147.

Dupont, E., Hanganu, I. L., Kilb, W., Hirsch, S., & Luhmann, H. J. (2006). Rapid developmental switch in the mechanisms driving early cortical columnar networks. *Nature, 439,* 79–83.

Edmondson, J. C., & Hatten, M. E. (1987). Glial-guided granule neuron migration in vitro: A high-resolution time-lapse video microscopic study. *Journal of Neuroscience, 7,* 1928–1934.

Edmondson, J. C., Liem, R. K., Kuster, J. E., & Hatten, M. E. (1988). Astrotactin: A novel neuronal surface antigen that mediates neuron-astroglial interactions in cerebellar microcultures. *Journal of Cell Biology, 106,* 505–517.

Eriksson, P. S., Perfilieva, E., Bjork-Eriksson, T., Alborn, A. M., Nordborg, C., Peterson, D. A., & Gage, F. H. (1998). Neurogenesis in the adult human hippocampus. *Nature Medicine, 4,* 1313–1317.

Etienne-Manneville, S., & Hall, A. (2001). Integrin-mediated activation of Cdc42 controls cell polarity in migrating astrocytes through PKCzeta. *Cell, 106,* 489–498.

Etienne-Manneville, S., & Hall, A. (2003). Cell polarity: Par6, aPKC and cytoskeletal crosstalk. *Current Opinion in Cell Biology, 15,* 67–72.

Farkas, L. M., & Huttner, W. B. (2008). The cell biology of neural stem and progenitor cells and its significance for their proliferation versus differentiation during mammalian brain development. *Current Opinion in Cell Biology, 20,* 707–715.

Faulkner, N. E., Dujardin, D. L., Tai, C. Y., Vaughan, K. T., O'Connell, C. B., Wang, Y., & Vallee, R. B. (2000). A role for the lissencephaly gene L1S1 in mitosis and cytoplasmic dynein function. *Nature Cell Biology, 2,* 784–791.

Feng, L., Hatten, M. E., & Heintz, N. (1994). Brain lipid-binding protein (BLBP): A novel signaling system in the developing mammalian CNS. *Neuron, 12,* 895–908.

Fishell, G., & Hatten, M. E. (1991). Astrotactin provides a receptor system for CNS neuronal migration. *Development, 113,* 755–765.

Fishell, G., Mason, C. A., & Hatten, M. E. (1993). Dispersion of neural progenitors within the germinal zones of the forebrain. *Nature, 362,* 636–638.

Gadisseux, J.F., & Evrard, P. (1985). Glial-neuronal relationship in the developing central nervous system. A histochemical-electron microscope study of radial glial cell particulate glycogen in normal and reeler mice and the human fetus. *Developmental Neuroscience, 7,* 12–32.

Gage, F. H. (2000). Mammalian neural stem cells. *Science, 287,* 1433–1438.

Gaiano, N., & Fishell, G. (2002). The role of notch in promoting glial and neural stem cell fates. *Annual Review of Neuroscience, 25,* 471–490.

Gammill, L., Gonzalez, C., Gu, C., & Bronner-Fraser, M. (2006). Guidance of trunk neural crest migration requires Neuropilin-2/Semaphorin3F signaling. *Development, 133,* 99–106.

Garcia-Moreno, F., Lopez-Mascaraque, L., & De Carlos, J. A. (2007). Origins and migratory routes of murine Cajal-Retzius cells. *Journal of Comparative Neurology, 500,* 419–432.

Gasser, U., & Hatten, M. (1990). Neuron-glia interactions of rat hippocampal cells in vitro: Glial-guided neuronal migration and neuronal regulation of glial differentiationsss. *Journal of Neuroscience, 10,* 1276–1285.

Gasser, U. E., & Hatten, M. E. (1990). Central nervous system neurons migrate on astroglial fibers from heterotypic brain regions in vitro. *Proceedings of the National Academy Sciences of the United States of America, 87,* 4543–4547.

Ghashghaei, H. T., Weber, J., Pevny, L., Schmid, R., Schwab, M. H., Lloyd, K. C., …, Anton, E. S. (2006). The role of neuregulin-ErbB4 interactions on the proliferation and organization of cells in the

subventricular zone. *Proceedings of the National Academy Sciences of the United States of America, 103,* 1930–1935.

Ghosh, A., & Shatz, C. J. (1992). Pathfinding and target selection by developing geniculocortical axons. *Journal of Neuroscience, 12,* 39–55.

Ghosh, A., & Shatz, C. J. (1993). A role for subplate neurons in the patterning of connections from thalamus to neocortex. *Development, 117,* 1031–1047.

Ghosh, A., Antonini, A., McConnell, S. K., & Shatz, C. J. (1990). Requirement for subplate neurons in the formation of thalamocortical connections. *Nature, 347,* 179–181.

Goffinet, A. M. (1984). Events governing organization of postmigratory neurons: Studies on brain development in normal and reeler mice. *Brain Research, 319,* 261–296.

Gomes, E. R., Jani, S., & Gundersen, G. G. (2005). Nuclear movement regulated by Cdc42, MRCK, myosin, and actin flow establishes MTOC polarization in migrating cells. *Cell, 121,* 451–463.

Gong, S., Zheng, C., Doughty, M. L., Losos, K., Didkovsky, N., Schambra, U. B., …, Heintz, N. (2003). A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature, 425,* 917–925.

Gotz, M., & Huttner, W. B. (2005). The cell biology of neurogenesis. *Nature Reviews Molecular Cell Biology, 6,* 777–788.

Govek, E. E., Hatten, M. E., & Van Aelst, L. (2011). The role of Rho GTPase proteins in CNS neuronal migration. *Developmental Neurobiology, 71*(6), 528–553.

Gregory, W. A., Edmonson, J. C., Hatten, M. E., & Mason, C. A. (1988). Cytology and neuron-glial apposition of migrating cerebellar granule cells in vitro. *Journal of Neuroscience, 8,* 1728–1738.

Hand, R., Bortone, D., Mattar, P., Nguyen, L., Heng, J. I., Guerrier, S., …, Polleux, F. (2005). Phosphorylation of Neurogenin2 specifies the migration properties and the dendritic morphology of pyramidal neurons in the neocortex. *Neuron, 48,* 45–62.

Harkmark, W. (1954). Cell migrations from the rhombic lip to the inferior olive the nucleus raphe and the pons A morphological and experimental investigation on chick embryos. *Journal of Comparative Neurology, 100,* 115–209.

Hatten, M. E. (1999). Central nervous system neuronal migration. *Annual Review of Neuroscience, 22,* 511–539.

Hatten, M. E. (2002). New directions in neuronal migration. *Science, 297,* 1660–1663.

Heasman, S. J., & Ridley, A. J. (2008). Mammalian Rho GTPases: New insights into their functions from in vivo studies. *Nature Reviews Molecular Cell Biology, 9,* 690–701.

Heng, J. I., Nguyen, L., Castro, D. S., Zimmer, C., Wildner, H., Armant, O., …, Guillemot, F. (2008). Neurogenin 2 controls cortical neuron migration through regulation of Rnd2. *Nature, 455,* 114–118.

Hinds, J. W., & Ruffett, T. L. (1971). Cell proliferation in the neural tube: An electron microscopic and Golgi analysis in the mouse cerebral vesicle. *Z Zellforsch Mikrosk Anat, 115,* 226–264.

His, W. (1886). Zur Geschichte des menschlichen Ruckenmarks und der Nervenwurzeln Ges Wissensch. BD 13 S, 477.

Kanold, P. O., Kara, P., Reid, R. C., & Shatz, C. J. (2003). Role of subplate neurons in functional maturation of visual cortical columns. *Science, 301,* 521–525.

Konig, N., & Marty, R. (1981). Early neurogenesis and synaptogenesis in cerebral cortex. *Bibliotheca Anatomica, 19,* 152–160.

Konig, N., Roch, G., & Marty, R. (1975). The onset of synaptogenesis in rat temporal cortex. *Anatomy and Embryology, 148,* 73–87.

Kostovic, I., & Rakic, P. (1980). Cytology and time of origin of interstitial neurons in the white matter in infant and adult human and monkey telencephalon. *Journal of Neurocytology, 9,* 219–242.

Kriegstein, A. R., & Gotz, M. (2003). Radial glia diversity: A matter of cell fate. *Glia, 43,* 37–43.

Kriegstein, A. R., & Noctor, S. C. (2004). Patterns of neuronal migration in the embryonic cortex. *Trends in Neurosciences, 27,* 392–399.

Kriegstein, A. R., & Parnavelas, J. G. (2006). Progress in corticogenesis. *Cerebral Cortex, 16,* i1–i2.

Krull, C. E., Lansford, R., Gale, N. W., Marcelle, C., Collazo, A., Yancopoulos, G., …, Bronner-Fraser, M. (1997). Interactions between Eph-related receptors and ligands confer rostrocaudal polarity to trunk neural crest migration. *Current Biology, 7,* 571–580.

Landsberg, R. L., Awatramani, R. B., Hunter, N. L., Farago, A. F., DiPetrantioio, H. J., & Dymecki, S. M. (2005). Hindbrain rhombic lip is comprised of disrete progenitor cell populations allocated by Pax6. *Neuron, 48,* 933–947.

LeDouarin, N. M. (1982). *The neural crest*. New York, NY: Cambridge University Press.

LeDouarin, N. M., & Kalcheim, C. (1999). *The neural crest*. New York, NY: Cambridge University Press.

Lein, E. S., Finney, E. M., McQuillen, P. S., & Shatz, C. J. (1999). Subplate neuron ablation alters neurotrophin expression and ocular dominance column formation. *Proceedings of the National Academy Sciences of the United States of America, 96,* 13491–13495.

Lois, C., & Alvarez-Buylla, A. (1994). Long-distance neuronal migration in the adult mammalian brain. *Science, 264,* 1145–1148.

Lois, C., Garcia-Verdugo, J. -M., & Alvarez-Buylla, A. (1996). Chain migration of neuronal precursors. *Science, 271,* 978–981.

Louvi, A., Yoshida, M., & Grove, E. A. (2007). The derivatives of the Wnt3a lineage in the central nervous system. *The Journal of Comparative Neurology, 504,* 550–569.

Luskin, M. B. (1998). Neuroblasts of the postnatal mammalian forebrain: their phenotype and fate. *Journal of Neurobiology, 36,* 221–233.

Machold, R., & Fishell, G. (2005). Math1 is expressed in temporally discrete pools of cerebellar rhombic-lip neural progenitors. *Neuron, 48,* 17–24.

Marin-Padilla, M. (1998). Cajal-Retzius cells and the development of the neocortex. *Trends in Neurosciences, 21,* 64–71.

McConnell, S. K., Ghosh, A., & Shatz, C. J. (1994). Subplate pioneers and the formation of descending connections from cerebral cortex. *Journal of Neuroscience, 14,* 1892–1907.

Mesngon, M. T., Tarricone, C., Hebbar, S., Guillotte, A. M., Schmitt, E. W., Lanier, L., …, Smith, D. S. (2006). Regulation of cytoplasmic dynein ATPase by LIS1. *Journal of Neuroscience, 26,* 2132–2139.

Meyer, G. (2007). Genetic control of neuronal migrations in human cortical development. *Advances in Anatomy, Embryology, and Cell Biology, 189,* 1 p preceding 1, 1–111.

Meyer, G., Soria, J. M., Martinez-Galan, J. R., Martin-Clemente, B., & Fairen, A. (1998). Different origins and developmental histories of transient neurons in the marginal zone of the fetal and neonatal rat cortex. *Journal of Comparative Neurology, 397,* 493–518.

Molliver, M. E., Kostovic, I., & van der Loos, H. (1973). The development of synapses in cerebral cortex of the human fetus. *Brain Research, 50,* 403–407.

Morales, D., & Hatten, M. E. (2006). Molecular markers of neuronal progenitors in the embryonic cerebellar anlage. *Journal of Neuroscience, 26,* 12226–12236.

Murase, S., & Horwitz, A. F. (2002). Deleted in colorectal carcinoma and differentially expressed integrins mediate the directional migration of neural precursors in the rostral migratory stream. *Journal of Neuroscience, 22,* 3568–3579.

Nakagawa, S., & Takeichi, M. (1998). Neural crest emigration from the neural tube depends on regulated cadherin expression. *Development, 125,* 2963–2971.

Nguyen-Ba-Charvet, K. T., Picard-Riera, N., Tessier-Lavigne, M., Baron-Van Evercooren, A., Sotelo, C., & Chedotal, A. (2004). Multiple roles for slits in the control of cell migration in the rostral migratory stream. *Journal of Neuroscience, 24,* 1497–1506.

Noctor, S. C., Flint, A. C., Weissman, T. A., Wong, W. S., Clinton, B. K., & Kriegstein, A. R. (2002). Dividing precursor cells of the embryonic cortical ventricular zone have morphological and molecular characteristics of radial glia. *Journal of Neuroscience, 22,* 3161–3173.

Noden, D. M. (1975). An analysis of the migratory behavior of avian cephalic neural crest cells. *Developmental Biology, 42,* 106–130.

Nowakowski, R. S., & Rakic, P. (1979). The mode of migration of neurons to the hippocampus: A Golgi and electron microscopic analysis in foetal rhesus monkey. *Journal of Neurocytology, 8,* 697–718.

Orioli, D., Henkemeyer, M., Lemke, G., Klein, R. & Pawson, T. (1996). Sek4 and Nuk receptors cooperate in guidance of commissural axons and in palate formation. *The EMBO Journal, 15*, 6035–6049.

Pacary, E., Heng, J., Azzarelli, R., Riou, P., Castro, D., Lbel-Potter, M., …, Guillemot, F. (2011). Proneural transcription factors regulate different steps of cortical neuron migration through Rnd-mediated inhibion of RhoAsignaling. *Neuron, 69,* 1069–1084.

Paratcha, G., Ibanez, C. F., & Ledda, F. (2006). GDNF is a chemoattractant factor for neuronal precursor cells in the rostral migratory stream. *Molecular and Cellular Neuroscience, 31,* 505–514.

Parnavelas, J. G., Barfield, J. A., Franke, E., & Luskin, M. B. (1991). Separate progenitor cells give rise to pyramidal and nonpyramidal neurons in the rat telencephalon. *Cerebral Cortex, 1,* 463–468.

Pencea, V., & Luskin, M. B. (2003). Prenatal development of the rodent rostral migratory stream. *Journal of Comparative Neurology, 463,* 402–418.

Petersen, P. H., Zou, K., Krauss, S., & Zhong, W. (2004). Continuing role for mouse Numb and Numbl in maintaining progenitor cells during cortical neurogenesis. *Nature Neuroscience, 7,* 803–811.

Petersen, P. H., Zou, K., Hwang, J. K., Jan, Y. N., & Zhong, W. (2002). Progenitor cell maintenance requires numb and numblike during mouse neurogenesis. *Nature, 419,* 929–934.

Pinto, L., & Gotz, M. (2007). Radial glial cell heterogeneity – the source of diverse progeny in the CNS. *Progress in Neurobiology, 83,* 2–23.

Polleux, F., Whitford, K. L., Dijkhuizen, P. A., Vitalis, T., & Ghosh, A. (2002). Control of cortical interneuron migration by neurotrophins and PI3-kinase signaling. *Development, 129,* 3147–3160.

Price, D. J., Kennedy, H., Dehay, C., Zhou, L., Mercier, M., Jossin, Y., …, Molnar, Z. (2006). The development of cortical connections. *The European Journal of Neuroscience, 23,* 910–920.

Rakic, P. (1972). Mode of cell migration to the superficial layers of fetal monkey cortex. *Journal of Comparative Neurology, 145,* 61–84.

Rakic, P. (1988). Defects of neuronal migration and the pathogenesis of cortical malformations. *Progress in Brain Research, 73,* 15–37.

Rakic, P. (2003). Elusive radial glial cells: historical and evolutionary perspective. *Glia, 43,* 19–32.

Rakic, S., & Zecevic, N. (2003). Emerging complexity of layer I in human cerebral cortex. *Cerebral Cortex, 13,* 1072–1083.

Reynolds, B. A., & Weiss, S. (1992). Generation of neurons and astrocytes from isolated cells of the adult mammalian central nervous system. *Science, 255,* 1707–1710.

Ridley, A. J., Schwartz, M. A., Burridge, K., Firtel, R. A., Ginsberg, M. H., Borisy, G., …, Horwitz, A. R. (2003). Cell migration: integrating signals from front to back. *Science, 302,* 1704–1709.

Rio, C., Rieff, H., Pelmin, Q., & Corfas, G. (1997). Neuregulin and erbB receptors play a critical role in neuronal migration. *Neuron, 19,* 39–50.

Rivas, R., & Hatten, M. (1995). Motility and cytoskeletal organization of migrating cerebellar granule neurons. *Journal of Neuroscience, 15,* 981–989.

Rodriguez, C. I., & Dymecki, S. J. (2000). Origin of the precerebellar system. *Neuron, 27,* 475–486.

Sauer, M. E., & Chittenden, A. C. (1959). Deoxyribonucleic acid content of cell nuclei in the neural tube of the chick embryo: evidence for intermitotic migration of nuclei. *Experimental Cell Research, 16,* 1–6.

Sauer, M. E., & Walker, B. E. (1959). Radioautographic study of interkinetic nuclear migration in the neural tube. *Proceeding of the Society for Experimental Biology and Medicine, 101,* 557–560.

Schneider, D., Gulacsi, A., & Hatten, M. E. (2011). Lrp12/Mig13a reveals changing patterns of preplate neuronal polarity during corticogenesis that are absent in reeler mutant mice. *Cerebral Cortex, 21,* 134–144.

Selleck, M. A. J., & Bronner-Fraser, M. (1995). Origins of the avian neural crest: The role of neural plate-epidermal interactions. *Development, 121,* 526–538.

Sharma, K., Korade, Z., & Frank, E. (1995). Late-migrating neuroepithelial cells from the spinal cord differentiate into sensory ganglion cells. *Neuron, 14,* 143–152.

Sidman, R. L. (1970). Autoradiographic methods and principles for study of the nervous system with thymidine-H3. In W. J. Nauta & S. O. E. Ebbesson (Eds.), *Contemporary research techniques of Neuroanatomy* (pp. 252–274). New York, NY: Springer-Verlag.

Sidman, R. L., & Rakic, P. (1973). Neuronal migration with special reference to developing human brain: A review. *Brain Research, 62,* 1–35.

Smith, D. S., Niethammer, M., Ayala, R., Zhou, Y., Gambello, M. J., Wynshaw-Boris, A., & Tsai, L. H. (2000). Regulation of cytoplasmic dynein behavior and mictotubule organization by mammalian Lis1. *Nature Cell Biology, 2,* 767–775.

Solecki, D. J., Model, L., Gaetz, J., Kapoor, T. M., & Hatten, M. E. (2004). Par6alpha signaling controls glial-guided neuronal migration. *Nature Neuroscience, 7,* 1195–1203.

Solecki, D. J., Trivedi, N., Govek, E. E., Kerekes, R. A., Gleason, S. S., & Hatten, M. E. (2009). Myosin II motors and F-actin dynamics drive the coordinated movement of the centrosome and soma during CNS glial-guided neuronal migration. *Neuron, 63,* 63–80.

Stemple, D. L., & Anderson, D. J. (1993). Lineage diversification of the neural crest: In vitro investigations. *Developmental Biology, 159,* 12–23.

Strobl-Mazzulla, P. H., Sauka-Spengler, T., & Bronner-Fraser, M. (2010). Histone demethylase JmjD2A regulates neural crest specification. *Developmental Cell, 19,* 460–468.

Sun, Y., Goderie, S. K., & Temple, S. (2005). Asymmetric distribution of EGFR receptor during mitosis generates diverse CNS progenitor cells. *Neuron, 45,* 873–886.

Tai, C. Y., Dujardin, D. L., Faulkner, N. E., & Vallee, R. B. (2002). Role of dynein, dynactin, and CLIP-170 interactions in LIS 1 kinetochore function. *Journal of Cell Biology, 156,* 959–968.

Tamaki, K., Fujimori, E., & Takauji, R. (1997). Origin and route of tangentially migrating neurons in the developing neocortical intermediate zone. *Journal of Neuroscience, 17,* 8313–8323.

Tan, S. -S., Faulkner-Jones, B., Breen, S. J., Walsh, M., Bertram, J. F., & Reese, B. E. (1995). Cell dispersion patterns in different cortical regions studied with an X-inactivated transgenic marker. *Development, 121,* 1029–1039.

Tan, S. S., Kalloniatis, M., Sturm, K., Tam, P., Reese, B., & Faulkner-Jones, B. (1998). Separate progenitors for radial and tangential cell dispersion during development of the cerebral neocortex. *Neuron, 21,* 295–304.

Tissir, F., & Goffinet, A. M. (2003). Reelin and brain development. *Nature Reviews Neuroscience, 4,* 496–505.

Tsai, J. W., Bremner, K. H., & Vallee, R. B. (2007). Dual subcellular roles for LIS1 and dynein in radial neuronal migration in live brain tissue. *Nature Neuroscience, 10,* 970–979.

Tsai, J. W., Chen, Y., Kriegstein, A. R., & Vallee, R. B. (2005). LIS1 RNA interference blocks neural stem cell division, morphogenesis, and motility at multiple stages. *Journal of Cell Biology, 170,* 935–945.

Umeshima, H., Hirano, T., & Kengaku, M. (2007). Microtubule-based nuclear movement occurs independently of centrosome positioning in migrating neurons. *Proceedings of the National Academy Sciences of the United States of America, 104,* 16182–16187.

Vicente-Manzanares, M., Webb, D. J., & Horwitz, A. R. (2005). Cell migration at a glance. *Journal of Cell Science, 118,* 4917–4919.

Walsh, C., & Cepko, C. L. (1992). Widespread dispersion of neuronal clones across functional regions of the cerebral cortex [see comments]. *Science, 255,* 434–440.

Wang, H. U., & Anderson, D. J. (1997). Roles of Eph family transmembrane ligands in repulsive guidance of trunk neural crest migration and motor axon outgrowth. *Neuron, 18,* 383–396.

Wang, V. Y., Rose, M. F., & Zoghbi, H. Y. (2005). Math1 expression redefines the rhombic lip derivatives and reveals novel lineages within the brainstem and cerebellum. *Neuron, 48,* 31–43.

Weston, J. A. (1963). A radiographic analysis of the migration and localization of trunk neural crest cells in the chick. *Developmental Biology, 6,* 279–310.

Wilson, P., Fryer, R. H., & Hatten, M. E. (2010). Astn2, A Novel member of the Astrotactin gene family, regulates the trafficking of ASTN1 during glial-guided neuronal migration. *Journal of Neuroscience, 30,* 8529–8540.

Wood, J., Martin, S., & Price, D. J. (1992). Evidence that the earliest generated cells of the murine cerebral cortex form a transient population in the subplate and marginal zone. *Developmental Brain Research, 66,* 137–140.

Wu, W., Wong, K., Chen, J., Jiang, Z., Dupuis, S., Wu, J. Y., & Rao, Y. (1999). Directional guidance of neuronal migration in the olfactory system by the protein Slit. *Nature, 400,* 331–336.

Xu, Q., Cobos, I., De La Cruz, E., Rubenstein, J. L., & Anderson, S. A. (2004). Origins of cortical interneuron subtypes. *Journal of Neuroscience, 24,* 2612–2622.

Yasuda, K., Momose, T., & Takahashi, Y. (2000). Applications of microelectroporation for studies of chick embryogenesis. *Development, Growth and Differentiation, 42*(3), 203–206.

Zecevic, N., & Rakic, P. (2001). Development of layer I neurons in the primate cerebral cortex. *Journal of Neuroscience, 21,* 5607–5619.

Zhang, J., Campbell, R. E., Ting, A. Y., & Tsien, R. Y. (2002). Creating new fluorescent probes for cell biology. *Nature Reviews Molecular Cell Biology, 3,* 906–918.

Zheng, C., Heintz, N., & Hatten, M. E. (1996). CNS gene encoding Astrotactin, which supports neuronal migration along glial fibers. *Science, 272,* 417–419.

Zhong, W., & Chia, W. (2008). Neurogenesis and asymmetric cell division. *Current Opinion in Neurobiology, 18,* 4–11.

Zhou, P., Porcionatto, M., Pilapil, M., Chen, Y., Choi, Y., Tolias, K. F., …, Segal, R. A. (2007). Polarized signaling endosomes coordinate BDNF-induced chemotaxis of cerebellar precursors. *Neuron, 55,* 53–68.

Suggested Readings

Bielas, S., Higginbotham, H., Koizumi, H., Tanaka, T., & Gleeson, J. G. (2004). Cortical neuronal migration mutants suggest separate but intersecting pathways. *Annual Review of Cell and Developmental Biology, 20,* 593–618.

Gao, L., & Macara, I. G. (2004). Isoforms of the polarity protein par6 have distinct functions. *Journal of Biological Chemistry, 279,* 41557–41562.

Gierdalski, M., Sardi, S. P., Corfas, G., & Juliano, S. L. (2005). Endogenous neuregulin restores radial glia in a (ferret) model of cortical dysplasia. *Journal of Neuroscience, 25,* 8498–8504.

Kanold, P. O., & Shatz, C. J. (2006). Subplate neurons regulate maturation of cortical inhibition and outcome of ocular dominance plasticity. *Neuron, 51,* 627–638.

Solecki, D. J., Liu, X. L., Tomoda, T., Fang, Y., & Hatten, M. E. (2001). Activated Notch2 signaling inhibits differentiation of cerebellar granule neuron precursors by maintaining proliferation. *Neuron, 31,* 557–568.

Marianne Bronner and Mary E. Hatten

This page intentionally left blank

CHAPTER

16

Growth Cones and Axon Pathfinding

Approximately 100 years ago, Ramón y Cajal described the trajectories taken by nerve cell processes as they extend toward their intermediate and final targets, and he inferred that multiple influences likely guide growing axons. Ramón y Cajal (1890) also observed for the first time the growing tips of axons and named them "growth cones." In an unparalleled feat of scientific conjecture based on morphological observations of fixed material, he described their behavior:

> From the functional point of view, one might say that the growth cone is like a club or battering ram endowed with exquisite chemical sensitivity, rapid ameboid movements, and a certain motive force allowing it to circumvent obstacles in its path, thus coursing between various cells until reaching its destination.

Decades later, Harrison and Speidel demonstrated using cell culture techniques the truth of Cajal's description of a highly motile, ameboid specialization at the tips of growing axons. Viewed in real time, growth cone shape changes very slowly, at a rate just detectable by an observer. Viewed with modern time-lapse techniques, however, the dynamism of their ever-changing morphology as they crawl forward is striking. These pioneering studies identified the growth cone as the key decision-making component in the elaboration of axonal pathways.

The growth cone at the distal tip of an axon extending in tissue culture is flattened into a thin fan-shaped sheet with many long, very thin spikes radiating forward (Fig. 16.1). The fan-shaped sheets are called lamellipodia, and the spikes are called filopodia, or microspikes. Many growth cones growing *in situ* have a similar appearance, but it is not unusual to observe spindle-shaped growth cones with tufts of forward-directed filopodia within axon bundles. More complex growth cone shapes are often characteristic of slowly extending growth cones choosing between possible routes of extension, whereas simpler morphologies are characteristic of rapidly extending growth cones coursing along permissive tracts.

GROWTH CONES ARE ACTIVELY GUIDED

Growth cones crawl forward as they elaborate the axons trailing behind them, and their extension is controlled by cues in their outside environment that ultimately direct them toward their appropriate targets. What is the nature of these guidance cues and how do they affect growth cone behavior? One view to account for the ability of the nervous system to wire itself, already implicit in Ramón y Cajal's writings, held that axonal growth is highly directed, with each class of axons navigating along a distinctive prescribed pathway to reach its target. For example, during neural development, spinal commissural axons extend seemingly unerringly toward the ventral-most region of the spinal cord, cross to the contralateral side, and then extend rostrally to targets in the brain (Fig. 16.2A). These neurons could be influenced by a variety of repulsive and attractive cues at different points along their trajectory, with each portion of the journey directed anew by distinct guidance mechanisms (Fig. 16.2B). An alternative view for wiring the nervous system postulated that axonal growth involves a random wandering of axons throughout the embryonic environment with connections appropriate for proper functioning of the nervous system somehow being maintained and reinforced at the expense of inappropriate connections.

A number of experiments provided evidence supporting directed axon growth, including the classic experiments of Sperry demonstrating highly specific targeting

© 2013 Elsevier Inc. All rights reserved.

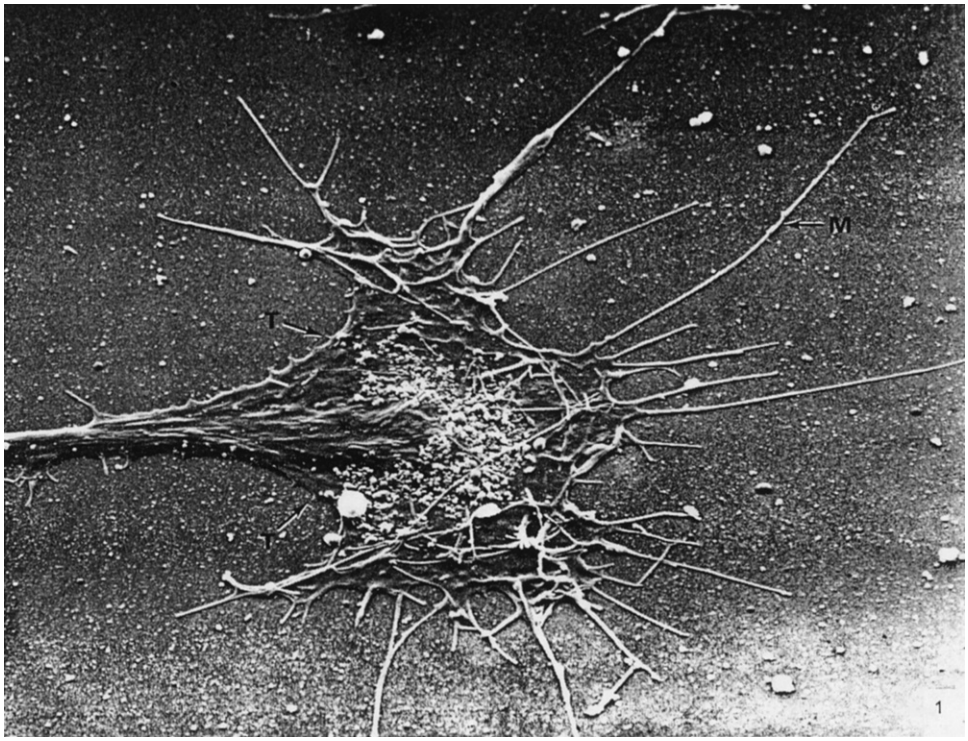


FIGURE 16.1 **Scanning electron micrograph of a growth cone in culture**. Growth cones extending on a flat surface are typically very thin, with broad lamellae and numerous filopodia. *Source: From Wessells and Nuttall (1978).*

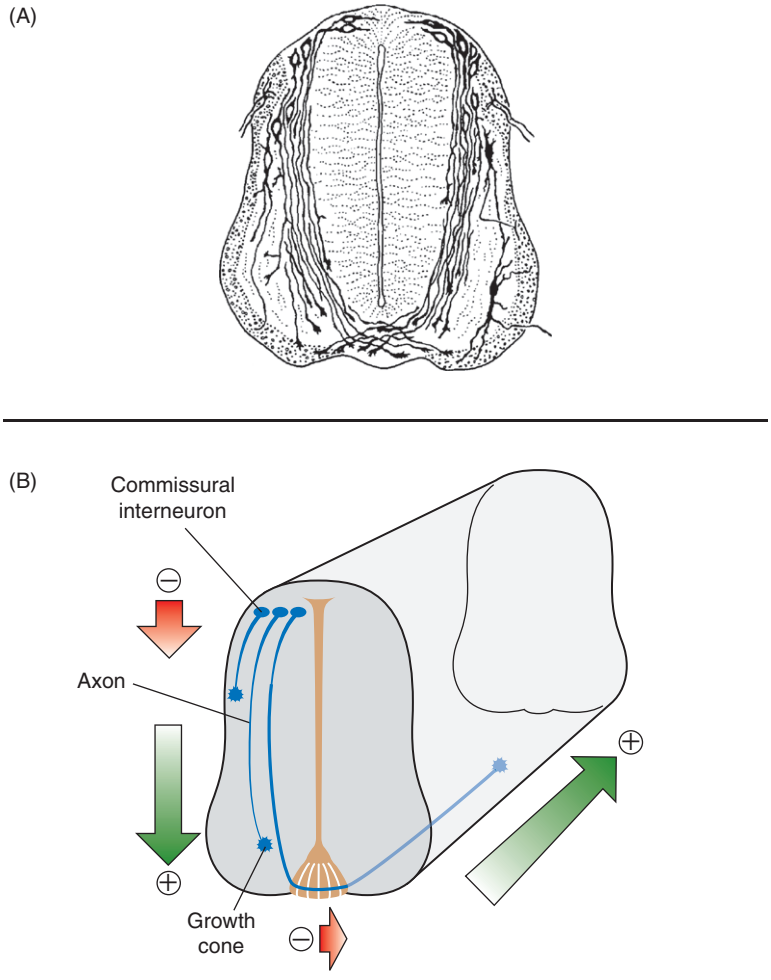


FIGURE 16.2 **Multiple guidance cues direct spinal cord commissural axons during neural development**. (A) This original drawing by Cajal illustrates neuronal pathways in the developing chick spinal cord, showing several commissural axons extending to the ventral spinal cord and crossing the floor plate. Note the distinct growth cone morphologies observed at different portions of the pathway, with growth cones at or crossing the midline exhibiting complex and expanded morphologies. From Ramón y Cajal (1890). (B) At least four classes of guidance cues are now known to instruct commissural axons during development: repellents from the dorsal spinal cord, initially directing commissural axons on a ventral course; attractive cues emanating from the ventral spinal cord guiding commissural axons toward the ventral spinal cord; repulsive cues in the ventral spinal cord, facilitating midline crossing and preventing recrossing; and attractants that guide postcrossing commissural axons anteriorly toward targets in the brain.

of regenerating retinal ganglion cell axons to the optic tectum (Sperry, 1963). The application of axonal tracing techniques has made it possible over the past 35 years to observe in multiple species a variety of different classes of axons en route to their targets. These studies show that the establishment of complex neural wiring patterns occurs through a combination of initial neuronal activity-independent guidance events and subsequent refinement of connections requiring electrical signaling among neurons (Goodman & Shatz, 1993). When projecting to the vicinity of their target field, axons grow along very stereotyped trajectories, making few projection errors. This growth is highly directed, and selection of the target cell is equally precise. A good example is the highly stereotyped and reproducible extension of individual axons arising from defined insect neurons as they grow within the central and peripheral nervous systems (Bastiani, du Lac, & Goodman, 1986; Caudy & Bentley, 1986; Raper, Bastiani, & Goodman, 1984). Neuronal ablation experiments demonstrate that individual neurons show selective affinity for certain axon bundles and in the absence of these targets fail to extend upon alternative neuronal, or nonneuronal, substrates (Fig. 16.3). Therefore, neurons appear programmed to connect with specific targets even before they contact them, an idea supported by the finding that axons of vertebrate motoneurons that have been surgically moved to abnormal nearby locations can reach their normal and appropriate targets even when they must traverse unusual routes to do so (Lance-Jones & Landmesser, 1980). Thus, guidance to target fields occurs prior to, and is independent of, neuronal circuit function. However, once many similar axons arrive at a target field containing numerous similar target cells, as occurs when groups of motor axons all innervate many muscle fibers in a single muscle or when thalamic neurons receiving input from both eyes establish overlapping axonal arborizations in the visual cortex, neuronal function does come into play. In these circumstances, individual axons initially arborize widely within the target field and contact many target cells, only later refining their pattern of connections in a process that depends on the precise patterns of electrical activity in these neurons and their target cells. The rest of this chapter focuses on the first step in the formation of connections: the directed growth of axons to target fields. The selection of target cells and the role of electrical activity in this process are discussed in Chapters 18, 20, and 22.

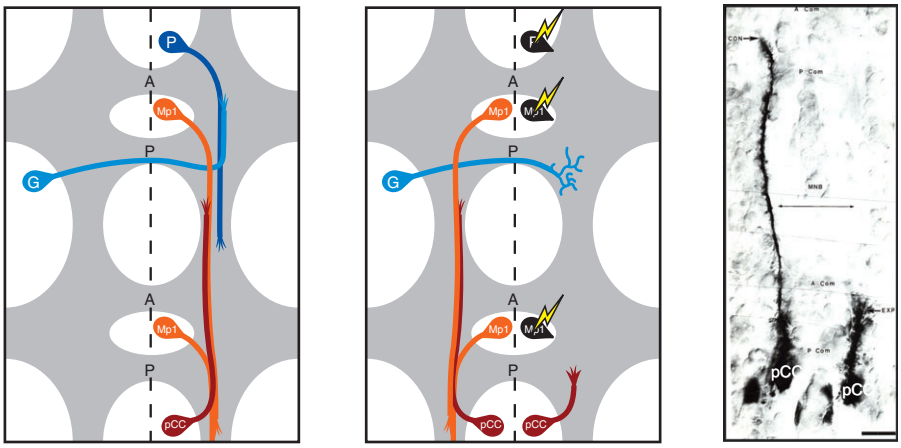


FIGURE 16.3 **Extending axons demonstrate exquisite selectivity in choosing their pathways during neural development.** (A) A simplified diagrammatic representation of the embryonic grasshopper central nervous system shows the establishment of two distinct axon pathways (or fascicles): the joining of the G neuron with the P trajectory, and the joining of the pCC neuron with the MP1 pathway. The G neuron extends across the CNS midline (dotted line) within the posterior commissure ("P") and selectively joins (or fasciculates) with axons from the P neurons (one of which is shown here), extending anteriorly along these descending P axons. The pCC neuron does not cross the midline and instead extends anteriorly along descending MP1 axons. (B) Ablation of neurons that contribute the P or MP1 pathways results in the inability of the G or pCC neurons, respectively, to extend anteriorly; these neurons do not fasciculate with numerous other axon pathways present very close by at this stage of development (not shown), and they instead stall at the point where they would have selectively fasciculated with their axonal targets. (C) Dye filling of pCC neurons on the experimental (right) and control (left) sides of a grasshopper embryo shows that when pCC does not encounter descending MP1 neurons, which are absent on the right side owing to ablation of these neurons, its growth cone stalls near the point where normally this contact would take place and no directed extension on alternative substrates occurs. *Source: From Bastiani et al. (1986) and Raper et al. (1984).*

GUIDANCE CUES FOR DEVELOPING AXONS

How do axons succeed in navigating through the embryonic environment to targets that in some cases can be many centimeters away? The trajectories of many axons appear to be broken up into short segments, each perhaps a few hundred micrometers long. The daunting task of reaching a distant target is then reduced to the simpler task of navigating each of these successive segments. This is illustrated by the projections of commissural neurons in the spinal cord, depicted in Figure 16.2. The pathway of these axons can be divided into discrete segments, each bounded by a specific cell, or group of cells, that marks the end of one segment and the beginning of the next. Such cells, forming an intermediate target for the axons, are sometimes termed "guidepost" cells. For example, commissural axons initially grow from the dorsal spinal cord to a set of guidepost cells at the ventral midline termed "floor plate cells," which serve as an important intermediate target. Commissural axons cross the midline at the floor plate, then turn and extend alongside the midline before eventually leaving the midline to seek out their ultimate targets. In mice, genetic ablation of floor plate cells results in profound misrouting of the axons when they reach the midline (Matise, Lustig, Sakurai, Grumet, & Joyner, 1999). Evidence for the existence of such guidepost cells harboring important guidance information has been obtained throughout the animal kingdom (Caudy & Bentley, 1986; McConnell, Ghosh, & Shatz, 1994).

The appreciation that axonal trajectories are formed in small segments pushes the question back one step: How do axons navigate each small segment of their trajectory? Axons appear to be guided along their appropriate trajectories by their responses to selectively distributed molecular signals within the developing embryo. Studies over the past three decades have led to the view that axon guidance involves the coordinate action of four types of cues: short-range (or local) cues and long-range cues, each of which can be either positive (attractive) or negative (repellent) (Fig. 16.4). The operations of short- and long-range guidance mechanisms and of attraction and repulsion are not mutually exclusive. Rather, axons may generally be guided over each individual segment of their trajectories by several different types of mechanisms acting in concert to ensure reproducible and high-fidelity guidance. Many experiments show that axons may navigate short segments by using several and, in some cases, even all four types of cues (Fig. 16.4): a repellent from behind the axons to "push," a corridor marked by a permissive local cue and bounded by an inhibitory local cue to "hem in" growth, and an attractant at the end of the corridor to "pull" (Tessier-Lavigne & Goodman, 1996). Push, pull, and hem—these forces working together can ensure accurate guidance.

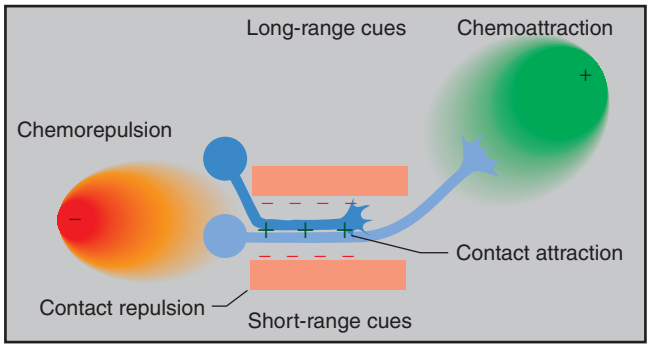
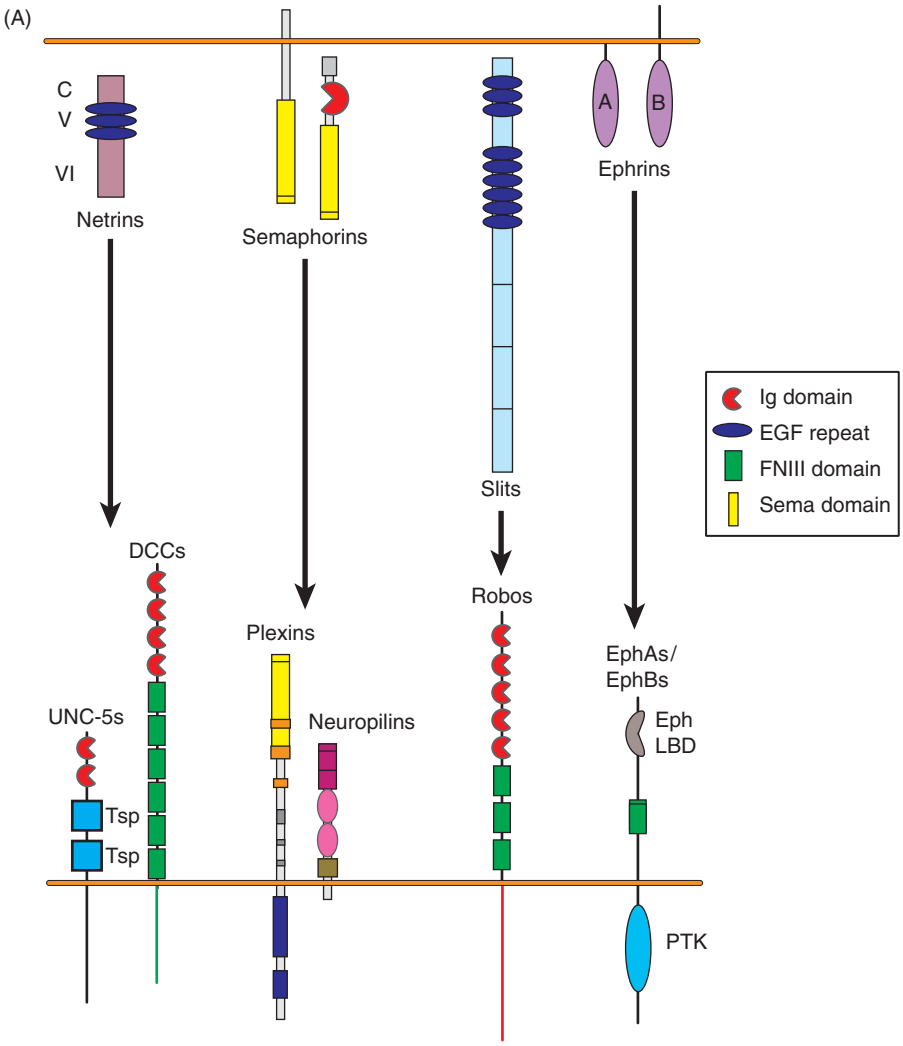


FIGURE 16.4 **Axons are guided by the simultaneous and coordinate actions of four types of guidance mechanisms:** contact attraction, chemoattraction, contact repulsion, and chemorepulsion. Individual growth cones might be "pushed" from behind by a chemorepellent, "pulled" from in front by a chemoattractant, and "hemmed in" by attractive and repulsive local cues (cell surface or extracellular matrix molecules). Push, pull, and hem: these forces act together to ensure accurate guidance. *Source: Adapted from Tessier-Lavigne and Goodman (1996).*

The identification of signaling molecules that function as guidance cues has depended primarily on three experimental approaches: (1) pairing biochemistry and *in vitro* tissue culture assays to detect proteins with either attractive or repellent properties, (2) using forward genetics to identify mutations that affect axon trajectories *in vivo*, or (3) using genetic and tissue culture approaches to characterize the functions of molecules with distributions or molecular structures that make them attractive candidate guidance cues.

Studies over the past decade and a half have, in this way, led to the identification of several prominent families of signaling molecules that function as instructive guidance cues (Kolodkin & Tessier-Lavigne, 2010). A first wave of studies in the 1990s identified four major families that have well-established and widespread roles in axon guidance: netrins, semaphorins, slits, and ephrins (Fig. 16.5A). These are sometimes referred to as the "classic" axon guidance molecules. More recently, two additional classes of molecules initially identified in other contexts were shown to function as guidance cues as well: proteins that are usually classified as morphogens (including members of the Bone Morphogenetic Protein, Hedgehog, Vascular Endothelial Growth Factor, and Wnt families), as well as some proteins initially identified as growth factors (Fig. 16.5B). The identification and characterization of these cues and their receptors have led to several important generalizations about guidance mechanisms (Dickson, 2002; Tessier-Lavigne & Goodman, 1996). First, guidance cues come



(B)

	Cues	Receptors	Sign of Response
Morphogens	Wnts	Frizzleds	+
		Derailed/Ryk	–
	HH	BOC + Smoothened	+
	Bmps	BMPR and BMPR-II	–
Growth Factors	GDNF	GFRα and Ret	+
	FGFs	FGFRs	+
	HGF	Met	+
	Neurotrophins	Trks	+

FIGURE 16.5 Multiple families of molecules serve as guidance cues. (A) Structures of the "canonical" axon guidance cues: netrins, semaphorins, slits, ephrins, and their receptors. Netrin receptors include members of the DCC and UNC5 family. There are seven classes of semaphorins (two invertebrate and five vertebrate). Of the ~20 mammalian semaphorins, 5 are secreted (class 3) and the others are divided into four families of transmembrane or GPI-linked proteins (represented here by a single diagram). Semaphorin receptors include plexins (of which nine are known in mammals, divided into classes A–D) and neuropilins (two in mammals). Class 3 semaphorin signaling is mediated by complexes of neuropilins (which serve as binding moieties) and plexins (signaling moieties). Slit receptors are members of the Roundabout (Robo) family. The receptors for Ephrins are members of the Eph family. Ephrin-As are GPI-anchored and predominantly bind EphAs, whereas Ephrin-Bs are transmembrane proteins and preferentially bind EphBs. *Adapted from Winberg et al. (1998) and Chisholm and Tessier-Lavigne (1999).* (B) Morphogens (members of the Wnt, Hedgehog, and BMP families) and various growth factors listed here also serve to guide neuronal processes. Known receptors functioning to mediate their effects in guidance are indicated.

in families that may, in some cases, comprise both diffusible members that can function in long-range axon guidance and nondiffusible members functioning at short range. Second, many guidance cues are multifunctional, attracting some axons, repelling others, and sometimes controlling axonal branching or arborization. Different axons may respond to the same cue differently because they express a distinct complement of surface receptors or contain particular components of the relevant signal transduction pathways. Third, many (although not all) cues are evolutionarily conserved between vertebrates and simpler invertebrate organisms, with species homologues often performing similar axon guidance roles. Higher vertebrates typically have many more members within a given family of guidance cues, and these cues are likely to have overlapping functions.

These different families of guidance cues are discussed briefly to illustrate features of wiring mechanisms. We

start by reviewing each of the four families of "classic" guidance cues before briefly describing the roles of morphogens and growth factors in guidance. We also focus specifically on members of the Immunoglobulin Gene (Ig) Superfamily, which have diverse roles as ligands and receptors in axons guidance.

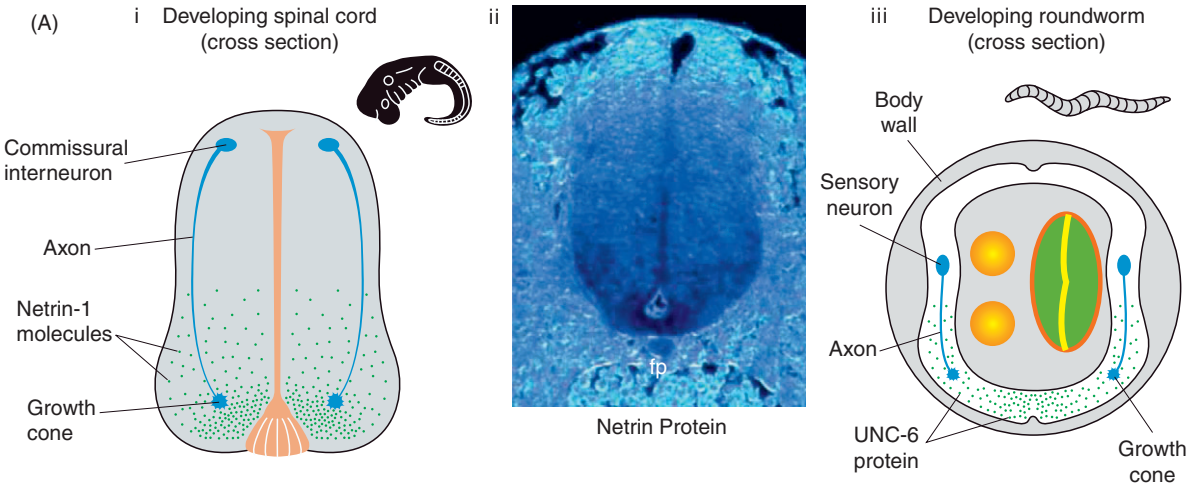
Netrins

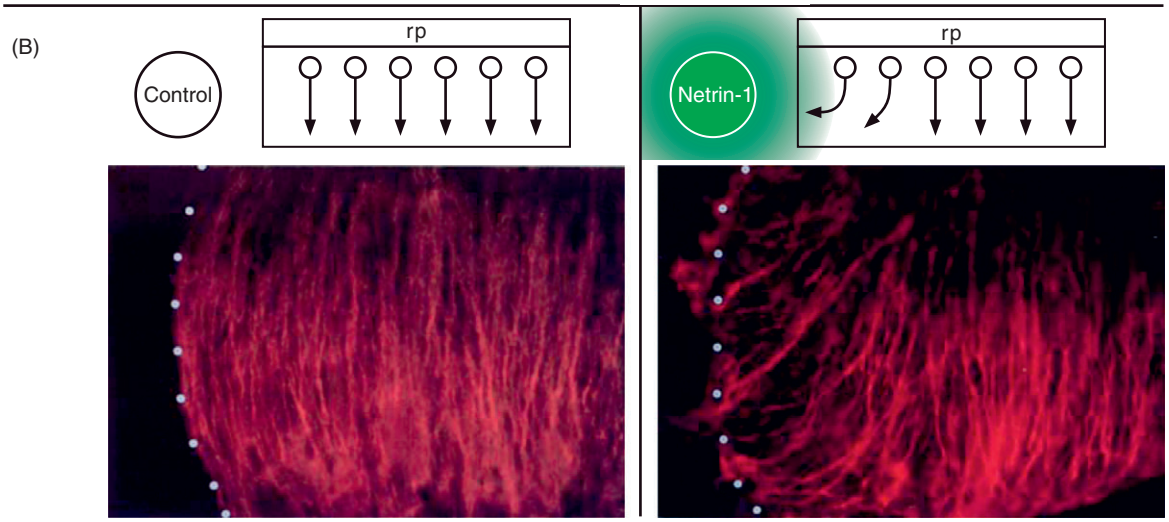
The netrin family (Fig. 16.5A) comprises about half a dozen members identified in vertebrates. The first vertebrate members were identified in studies on the guidance of commissural axons in the spinal cord. As described above, these axons project from the dorsal spinal cord to floor plate cells at the ventral midline (Fig. 16.6A). Experiments in tissue culture using microdissected pieces of rat spinal cord showed that floor plate cells secrete a diffusible outgrowth-promoting and chemoattractive activity for commissural axons, and netrins were identified as proteins that can mimic these activities (Serafini et al., 1994; Tessier-Lavigne, Placzek, Lumsden, Dodd, & Jessell, 1988). In the spinal cord, netrin is present in a dorsal-to-ventral increasing gradient (Figs. 16.6Ai and ii) and appears to provide guidance information to commissural axons, guiding them to the floor plate. In support of this view, netrin-1 can attract commissural axons from dorsal spinal cord pieces in culture, deflecting dorsal-ventral projections laterally when cells secreting recombinant netrin-1 are placed alongside the dorsal spinal cord pieces (Fig. 16.6B). Further, removal of netrin-1 function, which occurs in a *netrin-1* knockout mouse, results in significant misrouting of commissural axons, with many failing to reach the floor plate (Fig. 16.6C). Interestingly, a few axons do reach the floor plate even in the absence of netrin-1 (Fig. 16.6C), showing that one or more additional cues must collaborate to guide the axons. The collaboration of multiple guidance cues even over short segments of an axonal trajectory helps reduce the risk of guidance errors and contributes to the high degree of accuracy of wiring of the nervous system.

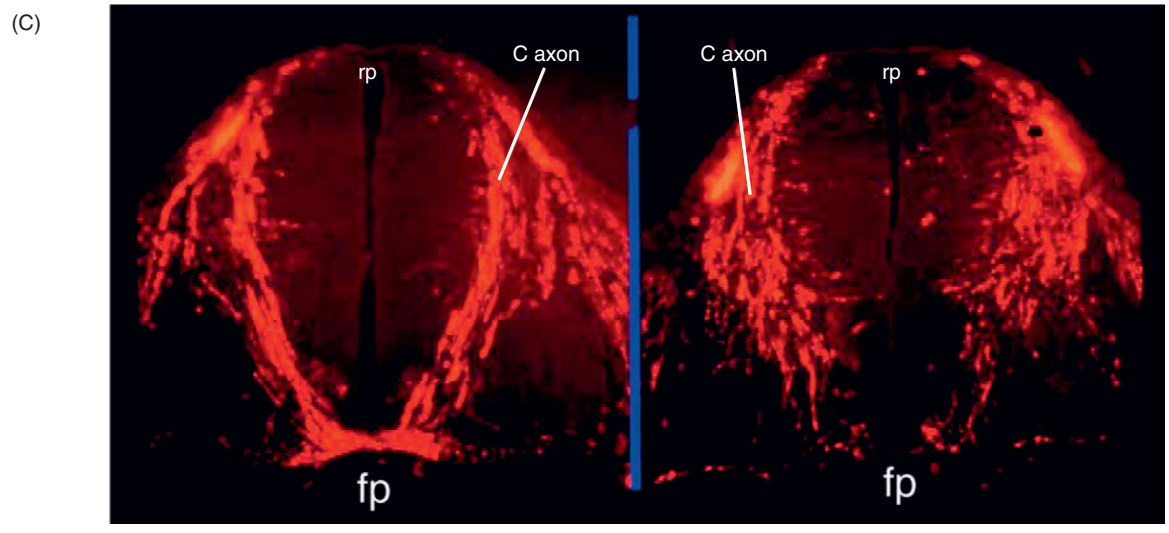
In addition to attracting commissural axons, netrins have been shown to function in attracting other subsets of axons to various targets in the retina and brain. Remarkably, netrin-1 also appears to function as a long-range repellent, providing a push from behind for a group of axons in the hind-brain that grow away from the midline, thus illustrating guidance cue bifunctionality.

Strikingly, netrins are vertebrate homologues of the UNC-6 protein of the nematode *Caenorhabditis elegans*, a protein similarly involved in both attracting some axons toward the nervous system midline and in repelling others away from it (Fig. 16.6Aiii) (Wadsworth, Bhatt, & Hedgecock, 1996). Netrin homologues are also expressed at the *Drosophila melanogaster* nervous system midline, where they contribute to attracting axons. In all of these organisms, netrin attraction is mediated by receptors of the DCC family, members of the immunoglobulin gene superfamily (Fig. 16.5) (Kolodkin & Tessier-Lavigne, 2010), and mutations in these receptors impair guidance to the midline in all species. This is true even in humans, since inherited loss-of-function mutations in the gene encoding DCC result in "mirror movements," a neurological syndrome in which voluntary movement of one extremity (e.g., the left arm) is accompanied by an involuntary mirror movement of the other (the right arm). This syndrome apparently results from miswiring of the nervous system, as corticospinal projections partially fail to reach and cross the midline at the pyramidal decussation so that each cerebral hemisphere is connected to both sides of the spinal cord, not simply the contralateral side as in unaffected humans (Srour et al., 2010). Netrin repulsion, in contrast, requires receptors of the UNC5 family, which are also members of the immunoglobulin gene superfamily (Figs. 16.5A) (Kolodkin & Tessier-Lavigne, 2010). In fact, as first shown in *C. elegans*, and illustrated for a vertebrate neuron in Figure 16.7, it is possible to convert

FIGURE 16.6 **Netrins are phylogenetically conserved axon attractants.** (A) Distribution and function of Netrins in axon guidance. (i) Diagram of a cross section through the embryonic vertebrate spinal cord depicting the role of Netrin-1 (here and in all subsequent schematic figures shown in green) in guiding commissural axons from the dorsal spinal cord to the ventral midline. (ii) A ventral to dorsal gradient of Netrin protein in the developing mouse spinal cord visualized using an immunohistochemical stain. *From Kennedy, Wang, Marshall, and Tessier-Lavigne (2006).* (iii) Diagram of a cross section through the developing nematode *C. elegans* illustrating the role of the Netrin/UNC-6 in guiding axons along a dorso-ventral trajectory. (B) Attractive effect of Netrin-1 on rodent commissural axons in an *in vitro* assay for chemoattraction. Pieces of the dorsal half of the spinal cord, viewed from the side, were placed in culture for 40 hours with either control COS cells or COS cells secreting Netrin-1. In control conditions (left), all the commissural axons grow from their cell bodies at the top (dorsal) side of the explants along a vertical (dorso-ventral) trajectory (shown in diagrammatic form at the top, and visualized within the spinal cord explants using an immunohistochemical axonal marker (red) at bottom). In the presence of cells secreting Netrin-1, however, the axons within ~150 μm of the cells (arrowheads) turn toward the Netrin source. Dots mark the boundary between COS cells and spinal cord tissue. In the presence of Netrin-1 axons can be seen invading the Netrin-secreting COS cell clump. *From Kennedy, Serafini, de la Torre, and Tessier-Lavigne (1994).* (C) Netrin-1 is required for guidance of commissural axons to the ventral midline. Shown are cross-sections through the spinal cord of embryonic day 11.5 mice. Left: wild-type (control); right: Netrin-1 knockout mouse. Axons are visualized with an axonal marker. Normally (left), commissural axons (c) extend along a smooth dorsal-to ventral trajectory to the floor plate (fp) at the ventral midline. In the absence of Netrin-1 (right), they extend normally in the dorsal half of the spinal cord, but when their trajectory through the ventral half is profoundly perturbed, with many projecting medially or laterally, only a very few axons reach the floor plate. *From Serafini et al. (1996).*


(A)
i Developing spinal cord (cross section)
Commissural interneuron
Axon
Netrin-1 molecules
Growth cone
ii
fp
Netrin Protein
iii Developing roundworm (cross section)
Body wall
Sensory neuron
Axon
UNC-6 protein
Growth cone


(B)
Control
rp
Netrin-1
rp


(C)
rp
C axon
fp
C axon
rp
fp
Wild Type
Netrin-1−/−

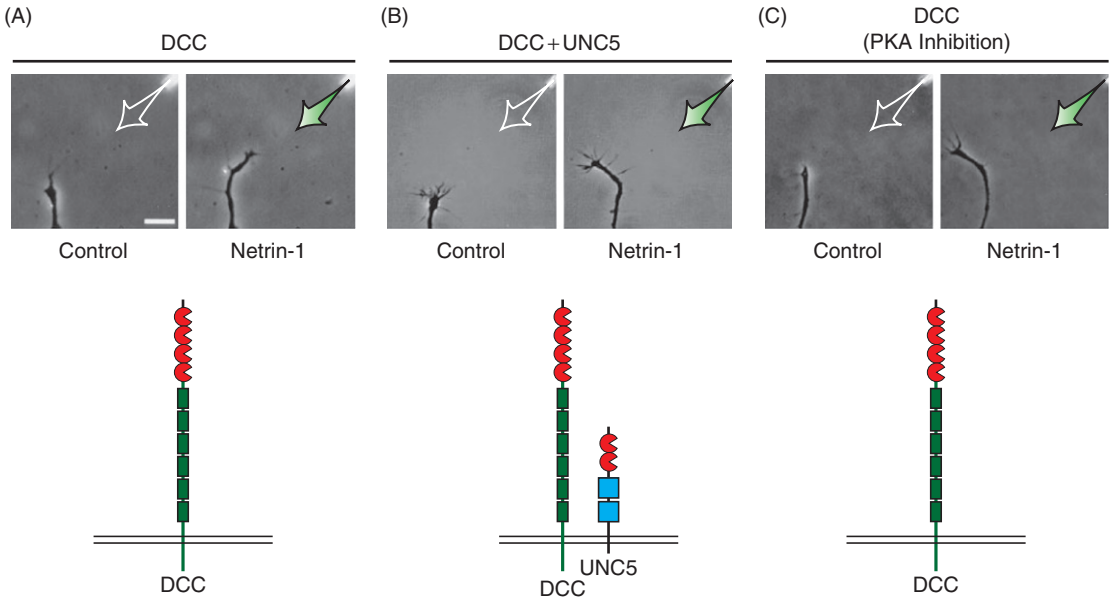


FIGURE 16.7 **Netrin bifunctionality depends upon both receptor composition and intracellular signaling.** (A) Single *Xenopus* spinal neurons in culture are attracted and steer along an increasing gradient of Netrin-1 originating from a micropipette (hollow arrow, upper right: green arrow, Netrin; white outlined arrow, control). These neurons express DCC on their surface. (B) If a *Xenopus* neuron expresses both DCC and UNC5 receptors, these proteins form a signaling complex which now interprets Netrin-1 as a repellent, and they steer down the gradient. (C) Inhibition of cAMP-mediated signaling using a protein kinase A (PKA) inhibitor also converts Netrin-1 attraction to repulsion, even though these neurons only express DCC. *From Ming et al. (1997) and Hong et al. (1999).*

netrin attraction, mediated by DCC, to repulsion by introducing an UNC5 family member into the neuron (Hamelin, Zhou, Su, Scott, & Culotti, 1993; Hong et al., 1999). These findings on netrins and their receptors vividly illustrate both guidance cue bifunctionality and the remarkable evolutionary conservation of axon guidance mechanisms.

The signaling mechanisms mediating attraction and repulsion are currently the subject of intense investigation (Bashaw & Klein, 2010). Several observations suggest that the signaling machinery within growth cones is organized to allow rapid switching from attractive to repulsive responses. For example, the responses of growth cones that are attracted to netrin-1 can be converted to repulsive responses through pharmacological manipulations that result in a reduction in cyclic AMP (cAMP) signaling (for instance, incubation with a membrane permeable inhibitor of the cAMP effector protein kinase A (PKA)) (Fig. 16.7) (Ming et al., 1997). How cAMP signaling modulates the netrin response is not fully understood, but it is clear that the growth cone turning machinery is organized to facilitate such switching.

Although netrins are capable of long-range attraction, the extent to which netrins can diffuse appears to be regulated so that in some circumstances they function as local cues, as shown in the nematode *C. elegans*, in the vertebrate retina, and at the *Drosophila* midline (Brankatschk & Dickson, 2006; Deiner et al., 1997; Wadsworth et al., 1996). Thus, like the distinction between attractive and repulsive cues, there is not a hard and fast distinction between local and long-range cues.

Semaphorins

The semaphorin family of proteins includes many secreted and membrane-associated proteins (Raper, 2000). The first known semaphorin was grasshopper Semaphorin 1a, a transmembrane protein required for correct pathfinding of pioneer axons in the developing limb (Kolodkin et al., 1992). The first vertebrate semaphorin, semaphorin 3A (Sema3A, originally called collapsin-1), was identified utilizing an *in vitro* axon repulsion assay to facilitate the biochemical purification of axonal repellents from brain extracts (Fig. 16.8A) (Luo, Raible, & Raper, 1993). There are approximately 20 distinct semaphorin family members in higher vertebrates, and all contain the family's signature semaphorin domain, a ~500 amino acid domain that is the key extracellular signaling domain of these proteins (Fig. 16.5A). Semaphorin biological specificity is in part determined by a relatively short stretch of amino acids within the semaphorin domain.

Most semaphorins have the capacity to act as potent inhibitory cues for neuronal or nonneuronal cells, and *in vitro* have been shown to repel very specific subsets of cultured axons. *In vivo* genetic analyses in invertebrates and vertebrates demonstrate conclusively that semaphorins serve as key axon repellents during neural

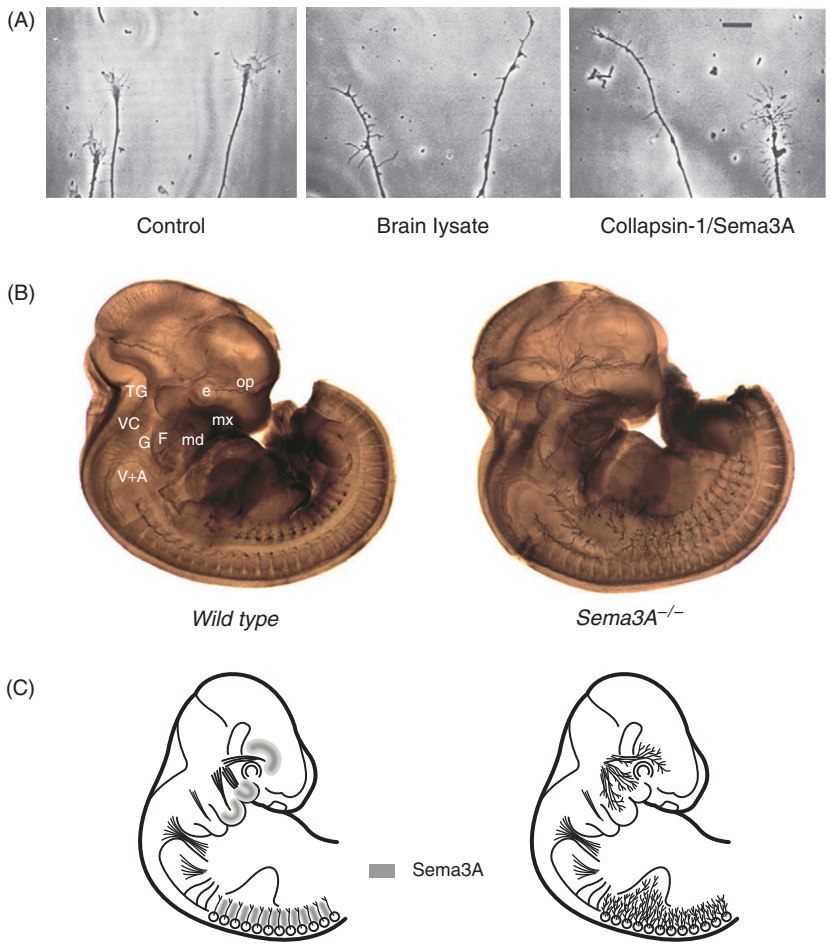


FIGURE 16.8 Secreted semaphorins are potent repellents. (A) Dorsal root ganglion neurons in culture exhibit a dramatic collapse when exposed to biochemically enriched fractions derived from chick brain lysates or to conditioned medium from tissue culture cells engineered to express recombinant Sema3A/collapsin-1. Medium from control non-transfected cells has no effect on DRG growth cone morphology. From Luo et al. (1993). (B) Mouse whole mount, embryonic day 11.5, wild-type and Sema3A mutant embryos immunostained for neurofilaments to illuminate spinal and cranial peripheral nerve projections. In Sema3A mutants, cranial nerves, including the trigeminal nerve (TG), are dramatically defasciculated, extend branches into ectopic locations, and overshoot their normal target regions. Spinal nerves, too, are defasciculated and extend into intersomitic regions they normally avoid (cranial nerves: ophthalmic, op; maxillary, mx; mandibular, md; facial, F, glossopharyngeal, G; vestibulocochlear, VC; vagus and accessory, V+A. Eye, e. Spinal nerves, arrow). Observations first reported by Taniguchi et al. (1997); whole-mount photographs courtesy of J. Merte and D. Ginty. (C) Schematic of whole mount embryos in B showing cranial and spinal nerve projections, and also Sema3A localization (gray).

development (Tran, Kolodkin, & Bharadwaj, 2007). For example, dramatic axon guidance defects are observed in *Sema3A* mutant mice (Fig. 16.8B) (Taniguchi et al., 1997). How does Sema3A-mediated repulsion contribute to normal axon guidance? Through various permissive cues expressed on axons or substrates upon which they extend promote axon outgrowth, repellents in the surrounding tissue constrain axon outgrowth—a process termed "surround repulsion." For example, Sema3A is expressed in tissues surrounding many peripheral nerves and acts as a repellent for sensory and motor axons. In *Sema3A* mutants, many peripheral nerves are defasciculated, extending and branching excessively into regions where normally Sema3A would inhibit their growth (Fig. 16.8B and C). Transmembrane semaphorins expressed on axons within fascicles also serve as repellents, but rather than providing "surround repulsion," they counter attractive axon-axon interactions, allowing axons to respond to other cues and defasciculate from main bundles in order to extend toward specific targets (Tran et al., 2007).

The major receptors for semaphorins are members of the plexin family, transmembrane proteins that include nine different receptors in higher vertebrates (Fig. 16.5A) (Tamagnone & Comoglio, 2000). Many transmembrane and some secreted semaphorins bind plexins directly. However, several secreted vertebrate semaphorins, including Sema3A, do not. Instead, these secreted semaphorins bind to the obligate co-receptors neuropilin-1 or neuropilin-2 that, together with a plexin family member, form an active receptor complex (Fig. 16.5A) (Tran et al., 2007). Plexin receptor activation initiates a series of intracellular signaling events that results in the local disassembly of growth cone cytoskeletal components and substrate attachments, ultimately producing inhibition of axon outgrowth or repulsive steering (Bashaw & Klein, 2010). Semaphorins also include both attractants and repellents, and the same semaphorin may under certain circumstances serve both functions. For example, the secreted semaphorin Sema3B in the mouse embryo functions as an attractant for one subset of axons in the developing anterior commissure tract in the brain and as a repellent for a different subset of axons in this same tract (Falk et al., 2005). Sema3A can act as both an attractant and a repellent for distinct neural processes extending from mouse cortical pyramidal neurons, and in a neuronal culture paradigm Sema3A repels

cortical pyramidal neuron axons but attracts dendrites from these same neurons (Polleux, Giger, Ginty, Kolodkin, & Ghosh, 1998; Polleux, Morrow, & Ghosh, 2000). Like netrin-1, Sema3A repulsive steering of isolated growths *in vitro* is also modulated by cyclic nucleotide signaling, but cGMP signaling plays a key role in switching Sema3A repulsion to attraction (Song et al., 1998).

Semaphorins facilitate the formation of central and peripheral axon pathways by regulating axon pathfinding and fasciculation, but semaphorins have also been implicated in targeting axons to the specific locations of their synaptic partners, in directing the pruning of exuberant projections, and in regulating the morphology of cortical pyramidal neuron dendrites (Pasterkamp & Giger, 2009).

Slits

Slits are large secreted proteins (Fig. 16.5A) that are also multifunctional. They were identified as axonal repellents through studies of chemorepellent factors in *Drosophila* and in vertebrates (Brose et al., 1999; Kidd, Bland, & Goodman, 1999; Li et al., 1999), and as stimulators of axon branching through studies of sensory axon arborization in vertebrates (Wang et al., 1999). A single Slit gene is found in *Drosophila* and in *C. elegans*, whereas there are three in mammals. Extensive studies in all three organisms demonstrate key roles for Slit-mediated repulsion in guidance of many axonal classes, including a variety of axons in the mammalian forebrain, as well as guidance at the midline of the spinal cord (discussed below). The branching activity of Slits has been observed in vertebrates (Ma & Tessier-Lavigne, 2007) and has been implicated in regulating dendritic branching (Whitford, Dijkhuizen, Polleux, & Ghosh, 2002).

In all species, the repulsive actions of Slit proteins are mediated by receptors of the Roundabout or Robo family (Kidd et al., 1998) (Fig. 16.5A), which, like the DCC and UNC5 families of netrin receptors, are also members of the immunoglobulin superfamily. Interestingly, whereas in the case of netrins different receptors mediate attraction and repulsion, for Slits it appears that both repulsion and branching are mediated by Robo family receptors, presumably activating distinct downstream effectors within the axons.

Ephrins

The fourth family of "classic" guidance cues consists of the ephrins, cell surface signaling molecules that play important roles in a large number of developmental events, including axon guidance (Klein, 2004). There are two subfamilies (Fig. 16.5A). Eight class A ephrins are tethered to the cell surface via GPI linkages, and six class B ephrins are transmembrane molecules. Ephrins must be clustered together to activate their receptors and do not appear to be active if released from the cell surface, so they are thought to function exclusively as short-range guidance cues. These ligands bind receptor tyrosine kinases of the Eph family. Class A ephrins interact with various degrees of selectivity with five class A Eph receptors, whereas class B ephrins interact with three class B Eph receptors (Fig. 16.5A). Ephrins have been shown to play an essential role in organizing topographic projections that connect, for example, retinal ganglion cells in the eye with their target cells in the appropriate portion of the optic tectum in lower vertebrates, or the lateral geniculate nucleus of the thalamus in higher vertebrates (see below). These mapping functions demonstrate the versatility of ephrins, which can function as attractants for some axons and repellents for others, and as either positive or negative regulators of axonal branching. In addition to topographic mapping, ephrins are implicated as short-range attractants and repellents in the guidance of a variety of central and peripheral axons, and they also play key roles in synapse development (see Chapter 18).

Morphogens and Growth Factors

The "classic" guidance cue families do not account for all axon guidance. Though certainly additional classes of guidance cues remain to be discovered, studies over the past decade reveal that various morphogens and growth factors have also been co-opted to guide axons (Fig. 16.5B). We are still learning how widely these molecules are used in this role.

Morphogens are proteins that function to specify cell fates at earlier stages of embryonic development; the term refers specifically to factors that can induce different fates in a dose-dependent fashion. Three key families of developmental morphogens have been implicated in axon guidance as well: members of the Hedgehog (Hh), Bone Morphogenetic Protein (BMP), and Wnt families. Most of these families were introduced in Chapter 14 in the discussion of cell fate determination. The evidence implicating these factors and their receptors in axon guidance is growing and is particularly well established in the vertebrate spinal cord. Here, members of all four families collaborate with classic guidance cues of the netrin and slit families (see below). Outside the spinal cord, evidence for guidance functions of BMP and Hedgehog proteins is still limited. Examples have been noted in *C. elegans* (involving the BMP family member UNC-129) and in the vertebrate retina (involving Sonic Hedgehog), but just how widely these factors function as guidance cues remains to be determined.

The evidence implicating Wnt family members in guidance is, however, quite extensive; they play key roles in axon guidance along the longitudinal (anterior-posterior) axis in *C. elegans*, *Drosophila*, and vertebrates, and they function in retinotectal topographic mapping (Schmitt et al., 2006). A list of these factors, and the receptors that mediate their guidance functions, is provided in Figure 16.5B.

Whereas morphogens that function at early stages of development have been co-opted to guide axons at later stages, various growth factors initially known for their functions at later stages of nervous system development—particularly in regulating the survival of neurons (Chapter 19)—also play earlier axon guidance roles. So far, these factors are all implicated as attractants: Hepatocyte Growth Factor (HGF), Fibroblast Growth Factors (FGFs), and Glial Cell-derived Neurotrophic Factor (GDNF) for subsets of motor axons, Neurotrophins (BDNF and NT3) for sensory axons, Neuregulins for thalamocortical axons, and Vascular Endothelial Growth Factor A (VEGFA) for commissural axons (Fig. 16.5B). In each case, the factors have been implicated in guidance of just one subset of axons and just in vertebrates. Thus, it remains an open question how widely growth factors are used to guide axons, and whether they can also function as repellents *in vivo*. There is, however, already considerable evidence that many of these factors function to stimulate branching of diverse classes of axons at their targets (for example, the neurotrophin NGF regulates branching of sensory axons at their targets, whereas BDNF regulates branching of thalamocortical axons in the cortex).

The following sections of this chapter explain how the guidance cues discussed above are deployed *in vivo* to affect accurate guidance. However, we first discuss one family of proteins, already mentioned above, that has diverse roles in regulating nervous system wiring, the Ig Superfamily proteins.

Ig Superfamily Proteins: Cell Adhesion Molecules, Guidance Cue Receptors, and Repellents

Initial work directed toward understanding nervous system connectivity in part focused on identifying factors that promote neuronal adhesion to substrates. In the 1980s, several cell adhesion molecules (CAMs) were identified, including a protein called Neural Cell Adhesion Molecule (NCAM) that was shown to mediate *in vitro* adhesive interactions in a homophilic (like-binds-like) fashion. NCAM contains a large extracellular domain structure that includes multiple repeats of an immunoglobulin (Ig)-C2 motif, and also several fibronectin type II repeats (Edelman, 1986). In cell culture NCAM can regulate axon-axon associations, but mice harboring mutations in *NCAM*, or in other vertebrate Ig superfamily adhesion molecules such as L1, do not exhibit guidance defects indicative of these proteins playing major instructive roles in nervous system wiring. However, the *Drosophila* cell adhesion molecule fasciclin II (FasII), the fly orthologue of NCAM, provides insight into how selective CAM expression can contribute to more subtle aspects of nervous system wiring—namely fasciculation (or bundling) of axons. FasII is expressed on a defined subset of fascicles in the developing insect CNS and PNS. Axons that contribute to FasII-expressing pathways still extend in approximately correct directions in *FasII* mutant fly embyros. However, ultrastructural analysis reveals in *FasII* mutants that axons that normally express FasII on their surface and maintain very close contact now fail to bundle together and instead proceed independently (Lin, Fetter, Kopczynski, Grenningloh, & Goodman, 1994).

A myriad of different Ig superfamily proteins is expressed in developing neurons, including CAMs and receptors for classic guidance cues (such as DCC, Robo, and Unc5; see above). However, a novel repulsive function for homophilic interactions involving the CAM Dscam (Downs Syndrome CAM) in *Drosophila* neural development have been observed (Schmucker et al., 2000; Zipursky & Sanes, 2010). Dscam is distinguished from all other Ig superfamily members by its genomic structure; alternative splicing of numerous exons has the capacity to result in the generation of close to 40,000 different Dscam isoforms. Homophilic Dscam interactions only occur between identical Dscam isoforms; a difference of even a few amino acids in otherwise identical isoforms precludes homophilic association. This astounding molecular diversity, coupled with expression of only a few isoforms in any one neuron, has the potential to allow individual neuronal processes, including axon branches or dendrites, to distinguish between self and nonself. This is because it is extremely unlikely that any two neurons will express the same complement of Dscam isoforms. Indeed, in the mushroom body of the *Drosophila* brain, isoform-specific Dscam interactions are required for axon branches from the same neuron to separate from one another and bifurcate at a single choice point (Zhan et al., 2004). Further, the elaboration of individual sensory neuron dendritic arborizations with nonoverlapping branches (called "tiling," which allows for an even distribution of sensory endings), also requires homophilic interactions between identical Dscam isoforms (Zipursky & Sanes, 2010). Therefore, Dscam isoform diversity allows neuronal processes emanating from a single neuron to be repelled from one another while simultaneously maintaining contact with processes from other neurons. Though in vertebrates Dscam is not extensively alternatively spliced, in the mammalian retina DSCAM plays a key role in

regulating self-avoidance among neurites of amacrine cell subtypes (Fuerst, Koizumi, Masland, & Burgess, 2008). Several other vertebrate neuronal proteins and families of proteins, including protocadherins, have the potential to provide extensive molecular diversity. Therefore, the employment of extreme protein isoform diversity in the service of distinguishing neuronal process uniqueness may be generally applicable across species (Zipursky & Sanes, 2010).

Summary

Several prominent classes of guidance cues have been identified: the four families of "classic guidance cues" (netrins, semaphorins, slits, ephrins) whose functions in guidance are widespread and well established, three families of morphogens (Hedgehogs, BMPs, and Wnts) whose functions are still being defined, a growing set of growth factors implicated in select attractive guidance events, and members of the Ig superfamily that, beyond their roles as receptors for other guidance cues, are implicated in axon fasciculation repulsion. These guidance cues can function in long-range or short-range axon guidance. Many guidance cues are multifunctional and can particpate in attraction, in repulsion, or in regulating branch formation. Furthermore, many cues are conserved evolutionarily, both in structure and in function, between vertebrates and invertebrates. In addition to these molecules, yet other molecules appear to contribute to guidance; these include a range of extracellular matrix components, transmembrane phosphatases, and cadherins.

GUIDANCE CUES AND THE CONTROL OF CYTOSKELETAL DYNAMICS

Guidance cues are signaling molecules that influence the cell biological mechanisms by which growth cones extend, turn, and retract. The forward-crawling motion of a growth cone depends on its own intrinsic motile mechanism interacting with a permissive outside environment, often in the context of instructive guidance cues. The extension and withdrawal of the leading edge and filopodia are an intrinsic property of a healthy growth cone. Once the leading edge has extended, an appropriate substratum on which it can attach and become stabilized must be present. Once attached, an inherent traction-generating mechanism within the growth cone causes tension to develop. Unattached or poorly attached processes are withdrawn, whereas tension exerted against attached processes helps draw the body of the growth cone forward. The continuous repeated cycling of extension, attachment, retraction of poorly attached processes, and tension generation produces net forward extension. In principle, guidance cues could act at any one of these steps to affect the direction in which growth cones extend.

Actin Cycle, Microtubules, and Axon Growth

Within growth cones the distribution of actin and microtubules is dramatically segregated, with microtubules primarily occupying the axon shaft and the central domain of the growth cone, and actin residing in the peripheral domain (Fig. 16.9). One way in which guidance cues can affect the direction of growth cone advance is by controlling the actin polymerization that helps drive protrusion of the leading edge of the growth cone. A dense meshwork of fibrillar actin (F-actin) is concentrated at the leading edge of the growth cone. New actin polymerization just behind the leading edge effectively helps push it forward, while, on average, F-actin is simultaneously depolymerized at an equal rate more centrally.

A second important component of growth cone motility is a continuous rearward flow of polymerized actin away from the leading edge toward the central domain. This actin flow can be visualized in cultured growth cones with the aid of a drug that blocks actin polymerization (Forscher & Smith, 1988). Actin monomers generated by depolymerization in the body of the growth cone are recycled to the front, polymerized again at the leading edge, and swept rearward, where they are depolymerized once again. This continuous rearward flow of polymerized actin is called "treadmilling," and it results from the action of myosin-based motors driving F-actin toward the growth cone central domain (Lin, Espreafico, Mooseker, & Forscher, 1996) (Fig. 16.10A). In the absence of adhesive contacts with the substratum, a steady state arises whereby the actin polymerization–depolymerization cycle and treadmilling do not contribute to growth cone advance. If polymerized F-actin is linked to a permissive substratum through cell surface receptors, however, growth cone actin dynamics advance the leading edge and withdraw the trailing edge (Fig. 16.10B). Cell surface receptors that bind permissive substratum molecules must in turn be linked to the cellular cytoskeleton. Indirect linkages between the cytoskeleton and members of the integrin, cadherin, and immunoglobulin families give the cytoskeleton traction upon the substratum and facilitate growth cone advance (Fig. 16.11B). This traction can be modulated by several signaling events, including the phosphorylation of specific cytoplasmic proteins (Dent, Gupton, & Gertler, 2010).

The linkage of F-actin to the substratum generates tension within the growth cone and causes processes that are poorly attached to the substratum to shrink or withdraw. When the retrograde movement of actin is impeded by a relatively strong attachment between the actin cytoskeleton and the surface substratum, not only

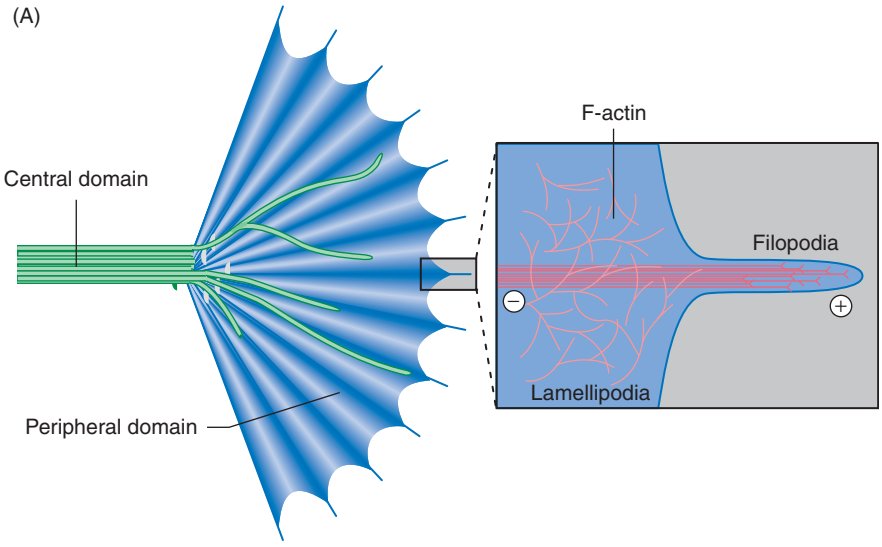


FIGURE 16.9 Distributions of microtubules and fibrillar actin in a neuronal growth cone. Microtubules and fibrillary actin are highly compartmentalized within growth cones. Microtubules (green) are a key structural component of the axon and though tightly bundled with the axon shaft, a dynamic subpopulation of microtubules actively explores the peripheral domain of the growth cone by extending along filopodia. All of the growing ends of microtubules are pointed toward the leading edge. In contrast, actin (red) is highly concentrated in the filopodia and in the leading edges of lamellae. Within filopodia, actin fibrils are oriented with their growing tips pointed distally. The same is true of many fibrils within lamellae, although many additional fibrils are oriented randomly and form a dense meshwork. *Modified from Lin, Thompson, and Forscher (1994).* (B) The organization of cytoplasmic domains and cytoskeletal components in an *Aplysia* bag cell growth cone. Fluorescent labeling of microtubules (green) and fibrillar actin (red) shows their segregation within the growth cone central and peripheral domains, respectively. Note the extension of a small number of microtubules into the actin-rich peripheral domain. *From Suter and Forscher (2000).*

is the leading edge of the growth cone advanced, but microtubules tend to be drawn forward. In the axon shaft and proximal region of the growth cone's central domain, microtubule polymers are bundled and relatively stable. However, toward the peripheral region a highly dynamic population of microtubule polymers extends individually toward the leading edge along filopodia. The strong attachment of individual filopodia to appropriate cellular targets induces additional localized actin accumulation in the contacting process, and these processes are preferentially invaded by dynamic microtubules. In this way filopodia act as scouts for less advanced portions of the growth cone: either seizing hold of permissive substrata or avoiding nonpermissive or inhibitory substrates, thereby reorienting subsequent process extension to define the location of the newly generated trailing axon shaft.

Particular representatives of several proteins families provide permissive substrata for growth cone extension. Among these are extracellular matrix (ECM) molecules such as laminin-1 and fibronectin, and also specialized cell surface molecules from the Ig superfamily and cadherin families. These molecules are bound by specific cell surface molecules on the growth cone, including

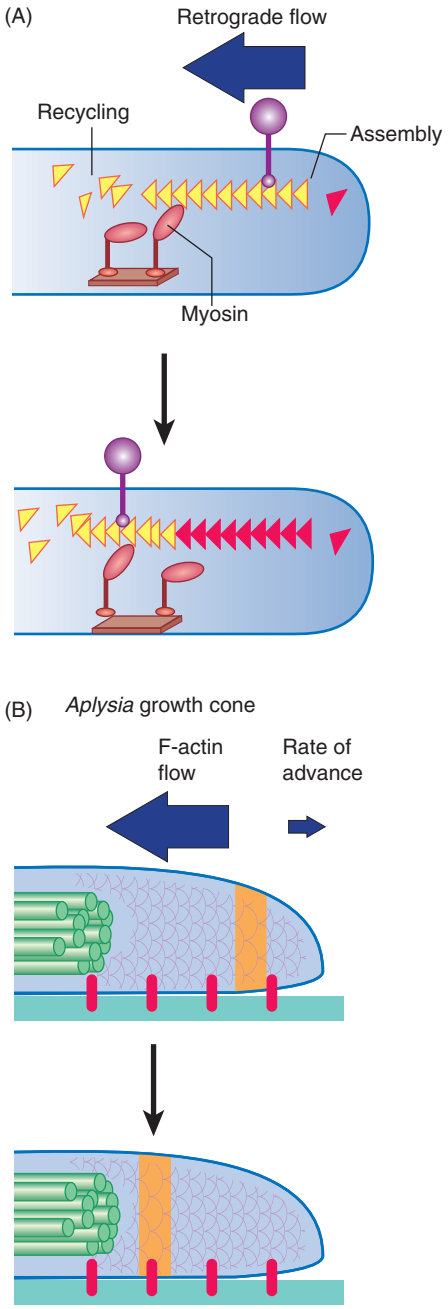


FIGURE 16.10 **Actin dynamics regulate growth cone advance.** (A) A schematic representation of a growth cone filopodium, depicting polarized assembly and disassembly of actin monomers at the distal (red triangles) and proximal (yellow triangles) ends of a fibrillar actin filament. Retrograde flow (large arrow) of fibrillar actin polymers occurs through the action of nonmuscle myosin. This can be visualized by the proximal translocation of a particle that is linked to the actin cytoskeleton (sphere), which fails to occur if myosin function is inhibited. *Modified from Lin et al. (1996).* (B) Linkage of the actin cytoskeleton to a permissive surface is required for forward advance. Actin polymerized at the leading edge of the growth cone (right) is swept toward the rear. If the actin meshwork is not linked to cell surface receptors that bind permissive molecules on adjacent cell surfaces, the actin cycles from front to rear but does not advance the growth cone. However, when the actin meshwork becomes attached to these receptors and thereby linked to the substrate, the meshwork remains in place and newly polymerized actin helps advance the leading edge. *Modified from Lin et al. (1994).*

members of the integrin, Ig, and cadherin superfamilies. While these proteins provide potent permissive substrates for cultured neurons, it is still unknown whether they or other classes of molecules are the major permissive factors for axon growth *in vivo*. At least two of the sets of "classic" guidance cues, the netrins and semaphorins, are known to include members that have a permissive role for axon growth *in vivo*. Interestingly, ECM components, including laminin-1 and select proteoglycans, can exert modulatory effects on classic guidance cue signaling, influencing the sign of the steering response (Höpker, Shewan, Tessier-Lavigne, Poo, & Holt, 1999; Kantor et al., 2004; Van Vactor, Wall, & Johnson, 2006).

Interactions Between the Cytoskeleton and Guidance Receptors

How do cues such as semaphorins, netrins, slits, and ephrins tap into these mechanisms for growth cone extension to affect guidance? Attractive and repulsive signals could promote the extension of the leading edge by increasing the local rate of actin polymerization in the preferred direction or by decreasing its rate of depolymerization in other directions. Guidance cues could also modulate the rate of myosin-mediated F-actin translocation, thereby stimulating or inhibiting filopodia and leading edge retraction. It is therefore reasonable to expect that many signaling pathways activated by guidance cues ultimately converge on actin dynamics. Studies on the receptor mechanisms for the guidance cues described above show that activation of these receptors often directly or indirectly alters the activity of small Rho family GTPases such as Rac and Rho, which are key regulators of actin dynamics and impact on cell motility in a large variety of cell types (Bashaw & Klein, 2010; Lalli & Hall, 2010). For example, the netrin receptor DCC activates Rac, whereas activation of certain semaphorin receptors results in activation of Rho through the action of an adaptor protein that associates with the cytoplasmic domain of Plexin receptors and stimulates the exchange of GDP for GTP on the Rho GTPase. Activation of both Eph and Robo receptors can also modulate Rho GTPases through the action of distinct adaptor proteins that associate with these receptors and either stimulate, or inactivate, Rho GTPases. In addition to modulating signaling cascades that regulate actin dynamics, guidance cue receptor activation can also directly regulate microtubule dynamics through activation of microtubule binding proteins that either promote or inhibit microtubule extension (Dent et al., 2010).

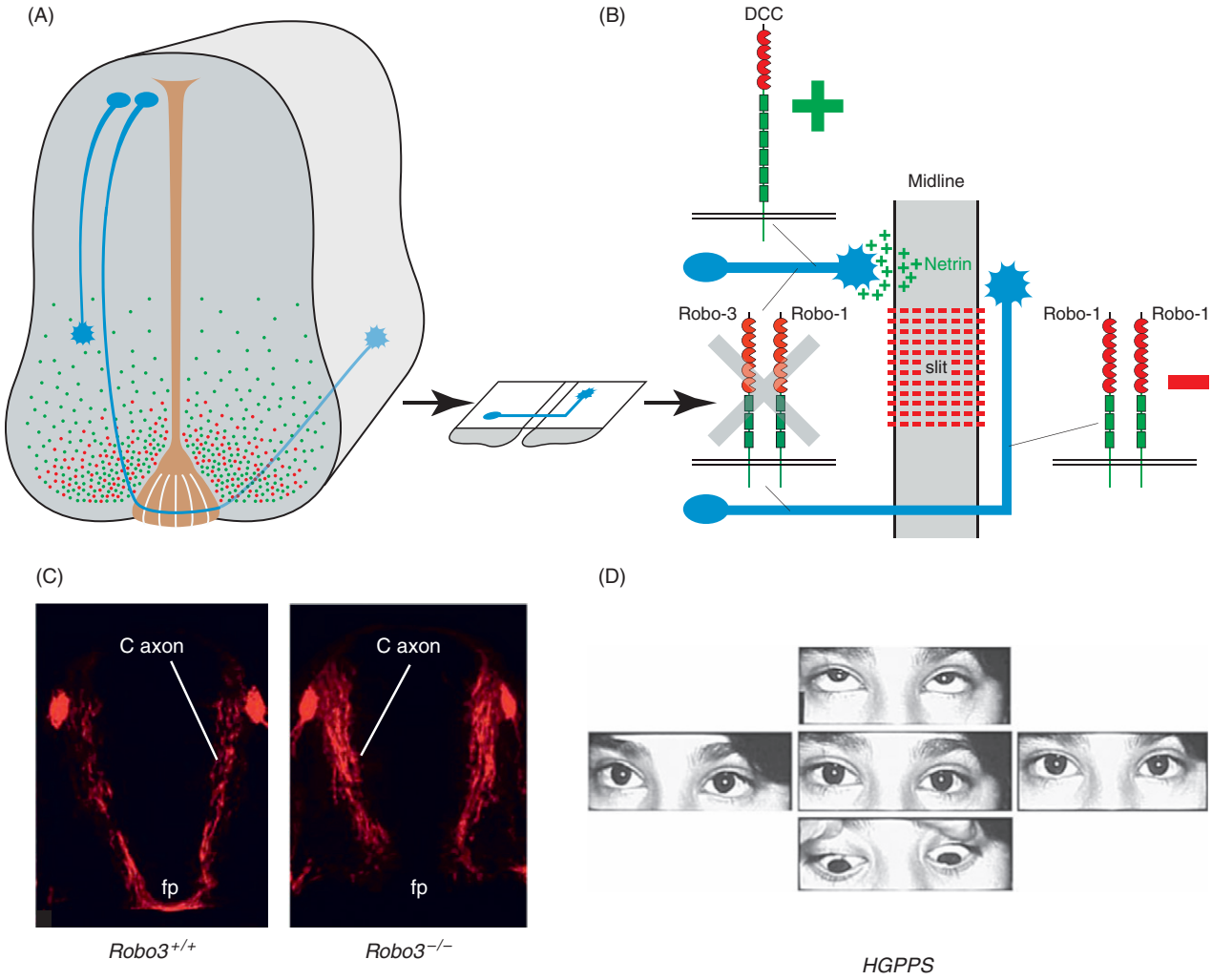


FIGURE 16.11 A switch from attraction to repulsion allows commissural axons to enter then leave the CNS midline. (A) Schematic of a cross section through the mammalian spinal cord (as in Figure 16.1), illustrating the trajectory of commissural axons from the dorsal spinal cord to, and across, the ventral midline. The axons cross the midline at the floor plate; upon exiting the floor plate, they make a sharp right-angle turn and project in a rostral (anterior) direction. A long-range gradient of Netrin-1 protein (green dots) and a shorter-range gradient of Slit protein (red dots) are shown. Right: Illustration of the "open-book" configuration, in which the spinal cord is opened at the dorsal midline and flattened, which helps to visualize the crossing and turning behavior of the axons at the midline. (B) As the axons project to the midline ("precrossing"), they are attracted by Netrin-1 acting on the attractive Netrin receptor DCC. They are insensitive to the repulsive action of Slit proteins despite expression of the repulsive Slit receptors (Robo1 and Robo2) because they also express Robo3, a Slit receptor that suppresses the activity of Robo-1 and Robo-2. This enables these axons to enter the midline upon reaching it. However, upon crossing the midline, commissural axons downregulate expression of Robo3 and increase expression of Robo1 and Robo2 (not shown), which allows them to sense the repulsive action of Slits and results in their expulsion from the midline. (C) In Robo3 mutant mice, the precrossing commissural axons are prematurely sensitive to Slits, which prevents them from entering the midline. Shown are cross sections through a normal E11.5 mouse embryo (left) and a stage-matched Robo3 mutant embryo (right), illustrating the inability of commissural axons in the mutant to cross the midline. *From Sabatier et al. (2004)*. (D) Human patients with Horizontal Gaze Palsy and Progressive Scoliosis (HGPPS) carry mutations in the human ROBO3 gene. They are capable of coordinated eye movements along the vertical axis (center), but not along the horizontal axis (left and right panels are attempts by this HGPPS patient to look left and right). The deficits in horizontal gaze result from a defect in neuronal connections across the midline in the hindbrain, which is presumed to arise from aberrant axon guidance during development, since a similar defect is observed in Robo3 mutant mice. *From Jen et al. (2004)*.

Summary

Growing axons utilize an intrinsic, actin-based treadmilling mechanism to extend filopodia and advance the leading edge of the growth cone. Growth cone extension requires an appropriate substratum for attachment and stabilization, which allows an inherent traction-generating mechanism within the growth cone to generate tension. This tension results in the withdrawal of poorly attached processes and the growth of processes that form attachments with the substratum. Net forward extension results from repeated cycles of extension, attachment, and the retraction of poorly attached processes. Cell surface

receptors that bind substratum molecules are linked indirectly to the cellular cytoskeleton, and these linkages can be modulated by the activation of signaling pathways. The binding of extracellular guidance cues to receptors on the growth cone surface modulates axon growth through several mechanisms, including regulation of actin polymerization, F-actin treadmilling, linkage between the growth cone cytoskeleton and the substrate, and modulation of microtubule dynamics. Many guidance receptors directly or indirectly modulate the activity of small Rho family GTPases, which are key regulators of actin polymerization.

GUIDANCE AT THE MIDLINE: CHANGING RESPONSES TO MULTIPLE CUES

How are guidance cues interpreted by growth cones *in vivo*, and how do they direct appropriate nervous system wiring? The guidance of an axon from its inception through its acquisition of a target is a complex process that can be broken down into a series of discrete decisions. Single guidance cues probably rarely, or never, act by themselves, and the integration of several simultaneously active cues is required for most, if not all, guidance decisions. This can be illustrated by interactions between the activities of many different guidance cues at the midline of the developing nervous systems in both invertebrates and vertebrates. These interactions have been elucidated through genetic, biochemical, and embryological studies in rodents, *Drosophila*, and *C. elegans*. Focusing on the mammalian spinal cord, we describe four important regulatory mechanisms that are required for reaching, crossing, and leaving the midline: initial attraction; switching on of repulsion; switching off of attraction; and prevention of premature repulsion. These mechanisms involve both classic guidance cues, illustrated in Figure 16.11, and morphogens, illustrated in Figure 16.12.

Initial Attraction

As mentioned earlier, in both vertebrates and invertebrates, the so-called commissural axons are attracted to the nervous system midline by members of the netrin family, which activate receptors of the DCC family (Kolodkin & Tessier-Lavigne, 2010; Tessier-Lavigne & Goodman, 1996). We also mentioned that netrins do not function alone in guiding commissural axons to the midline. In the mammalian spinal cord, the morphogen Sonic Hedgehog is also made by floor plate cells at the ventral midline and collaborates with netrin-1 to attract the axons to the midline (Fig. 16.12); recent work shows that the morphogen vascular endothelial growth factor A (VEGFA) serves a very similar function. In addition, BMP morphogens are made at the dorsal midline (by so-called "roof plate cells") and function as

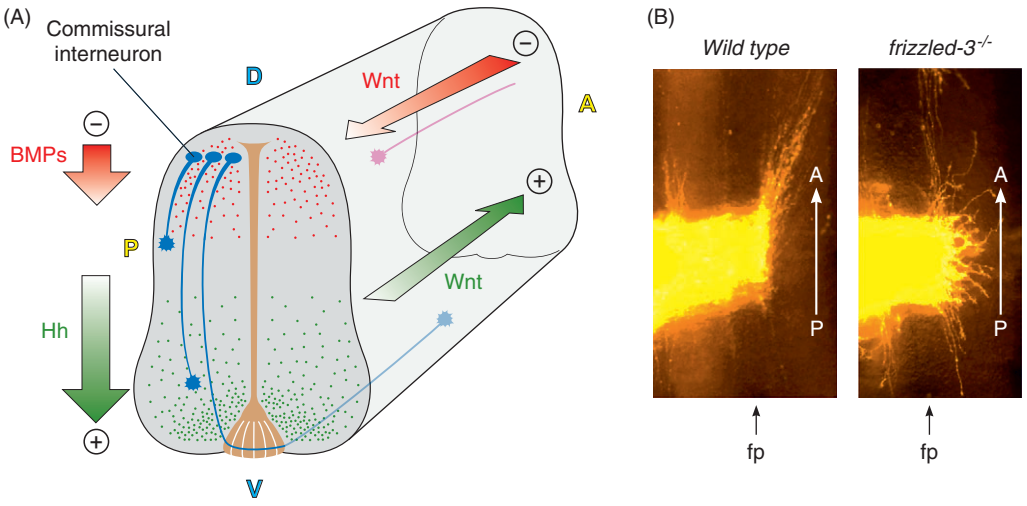


FIGURE 16.12 Morphogens help direct pre- and postcrossing commissural axon trajectories. (A) Diagram of the roles of morphogens in commissural axon guidance to the midline and beyond. BMP proteins made at the dorsal midline repel the axons from the dorsal spinal cord, providing a "push from behind," whereas Sonic Hedgehog (Shh) made by floor plate cells attracts the axons, providing a "pull from afar" in collaboration with Netrin-1 (see Fig. 16.11). After midline crossing, the axons turn in a rostral (anterior) direction in response to an attractive rostral-to-caudal decreasing gradient of Wnt protein. In contrast, cortical spinal tract (CST) axons extending caudally in the dorsal spinal cord are repelled by different Wnt family members. (B) The rostral Wnt signal is mediated by the Frizzled3 (Fz3) receptor. The trajectory of commissural axons at the midline is examined in the "open-book" configuration (Fig. 16.13A), with the axons visualized by labeling with the fluorescent dye DiI. In wild-type mice (left), commissural axons cross the midline, then turn rostrally (up), whereas in Frz3 knockout mice (right), the axons are disorganized after crossing (Lyuksyutova et al., 2009).

repellents for commissural axons, providing a "push from behind" to help set them out on their initial ventral trajectory. Thus, even over a relatively simply trajectory, multiple cues collaborate to ensure accurate guidance of axons.

Switching on Repulsion

The midline is not, however, the final destination for commissural axons: upon reaching it, axons cross the midline, then they turn at right angles and project alongside the midline to other levels of the embryo, and finally they leave the midline area altogether to reach their eventual targets (Fig. 16.11A). This behavior immediately raises a paradox: if the midline is such an attractive environment for the axons, how can they leave it? The answer is that midline cells, in addition to making attractive netrin proteins, also make repellents of the slit family (and, in vertebrates, of the semaphorin family) (Fig. 16.11A) (Kidd et al., 1999; Zou, Stoeckli, Chen, & Tessier-Lavigne, 2000). Axons can approach and then cross the midline a first time because they are initially insensitive to these repellents and thus respond only to netrin-mediated attraction. However, and quite remarkably, upon crossing the midline the axons become responsive to the repellents by increasing the expression or function of receptors for the repellents, including upregulating expression of the slit receptors of the Robo family on their surfaces (Fig. 16.11B) (Kidd et al., 1998; Zou et al., 2000). What causes this dramatic change in receptor expression and function is not yet known, but its net effect is to make the axons interpret the midline as a repulsive environment, which helps expel the axons from the midline and move on to the next leg of their trajectory.

Switching Off Attraction

The upregulation of responsiveness to midline repellents is only half of the equation, however: to be efficiently expelled from the midline, it would be desirable for the axons to stop being attracted. Indeed, the attractive response of these axons to netrins is switched off (silenced) as the axons cross the midline. There is evidence that the silencing of netrin attraction is caused by activation of the slit receptor Robo: activation of Robo by slit causes a specific region of its cytoplasmic domain to bind a specific region of the cytoplasmic domain of the netrin receptor DCC and to prevent it from transducing an attractive response to netrin (Stein & Tessier-Lavigne, 2001). Together, these two mechanisms (upregulation of a response to midline repellents and silencing of the attractive receptor) ensure that the growth cones perceive a once attractive environment, the midline, as unambiguously repulsive.

Preventing Premature Repulsion

For guidance at the midline to occur accurately, it is essential that the switch from attraction to repulsion be carefully choreographed. In particular, it is essential that growth cones upregulate responses to repellents only after crossing the midline, and not before, or the axons would not be able to enter the midline. It is perhaps not surprising, then, that there are mechanisms that ensure these axons do not become responsive to repellents too soon. One mechanism that regulates the switch, already mentioned, is that the Robo receptors are expressed only at low levels before crossing and are upregulated only after crossing. In *Drosophila*, this appears to involve regulated trafficking of the receptor to the neuronal cell surface under the control of a protein called Commissureless. However, the low level of Robo receptor that is present on the axons before crossing appears to be capable, nonetheless, of transducing a repulsive Slit signal that can prevent midline crossing and that must be silenced. In vertebrates, this is achieved by a more divergent member of the Robo family called Robo3 (also known as Rig1). Robo3 is expressed on commissural axons before and during crossing, and it silences the repulsive Slit signal mediated by other Robo receptors, allowing the axons to cross (Sabatier et al., 2004). Thus, in *Robo3* knockout mice, which lack Robo3 function, the axons can sense Slit repulsion before midline crossing. As a result, commissurals fail to cross the midline in these mutant mice, resulting in a dramatic "commissureless" phenotype (Fig. 16.11C).

Remarkably, a similar midline crossing defect is caused by mutations in the human *ROBO3* gene, found in patients with a syndrome known as "Horizontal Gaze Palsy with Progressive Scoliosis" (HGPPS) (Jen et al., 2004). These patients are unable to move their eyes in a coordinated fashion along the horizontal axis (horizontal conjugate gaze), a behavior that requires the function of commissural interneurons linking the two sides of the hindbrain (Fig. 16.11D). Imaging and evoked potential studies have shown that, as in *Robo3* knockout mice, there is a failure of midline crossing by axons in the hindbrains of these patients. Thus, the role of Robo3 proteins in allowing midline crossing is conserved in humans.

Guidance at the midline has also shed light on the mechanisms involved in longitudinal guidance (i.e., along the anterior-posterior axis). Upon crossing the midline, commissural axons turn either anteriorly or posteriorly; in particular, the earliest born commissural axons in the mammalian spinal cord turn in an anterior direction. This anterior turn appears to be directed by an anterior-to-posterior gradient of Wnt protein in the spinal cord, which the growth cones sense as an attractive cue (Fig. 16.12A) (Lyuksyutova et al., 2003). Transduction of the attractive Wnt signal involves a receptor called Frizzled 3 (Fz3), so that loss of Fr3 function in a

Fz3 knockout mouse causes the growth cones to become misrouted after midline crossing, unable as they are to sense the Wnt gradient (Fig. 16.12B). Anterior-posterior gradients of Wnt proteins provide a mechanism conserved between vertebrates and invertebrates to guide a variety of classes of axons along the longitudinal axis. Some axons, like the spinal commissural axons just described, respond to the Wnts with an attractive response and extend up the gradient. Others have repulsive responses and therefore grow downgradient. This has been shown for cortico-spinal tract axons, which navigate in the opposite direction to commissural axons (Fig. 16.12A) (Liu et al., 2005).

Summary

The study of the growth of commissural axons to and beyond the midline exemplifies how guidance of axons *in vivo* involves the simultaneous interpretation of multiple cues, including classic guidance cues and morphogens, and how dynamic changes in axon responsiveness enable axons to navigate successive intermediate targets. Before reaching the midline, commissural axons are guided through the combined actions of repellent BMPs, providing a "push from behind," and the attractive effects of netrin-1, Sonic Hedgehog, and VEGFA from the ventral midline. Just as important, these axons are insensitive to midline repellents, which allows them to enter and cross the midline; this is achieved in part by Robo3, expressed on the axons before crossing, which prevents them from sensing repellent midline Slits. Upon reaching the midline, the axons acquire responsiveness to repellent molecules, including Slits, through upregulation of repellent receptor function, including Robo receptors for midline Slits. They concomitantly *lose* responsiveness to the midline attractants. This coordinated switch from attraction to repulsion ensures that the axons can move on from the midline. Finally, the precise positioning of the axons in axon tracts after midline crossing involves a code of Robo proteins, and the anterior turn by commissural axons is directed by a Wnt gradient.

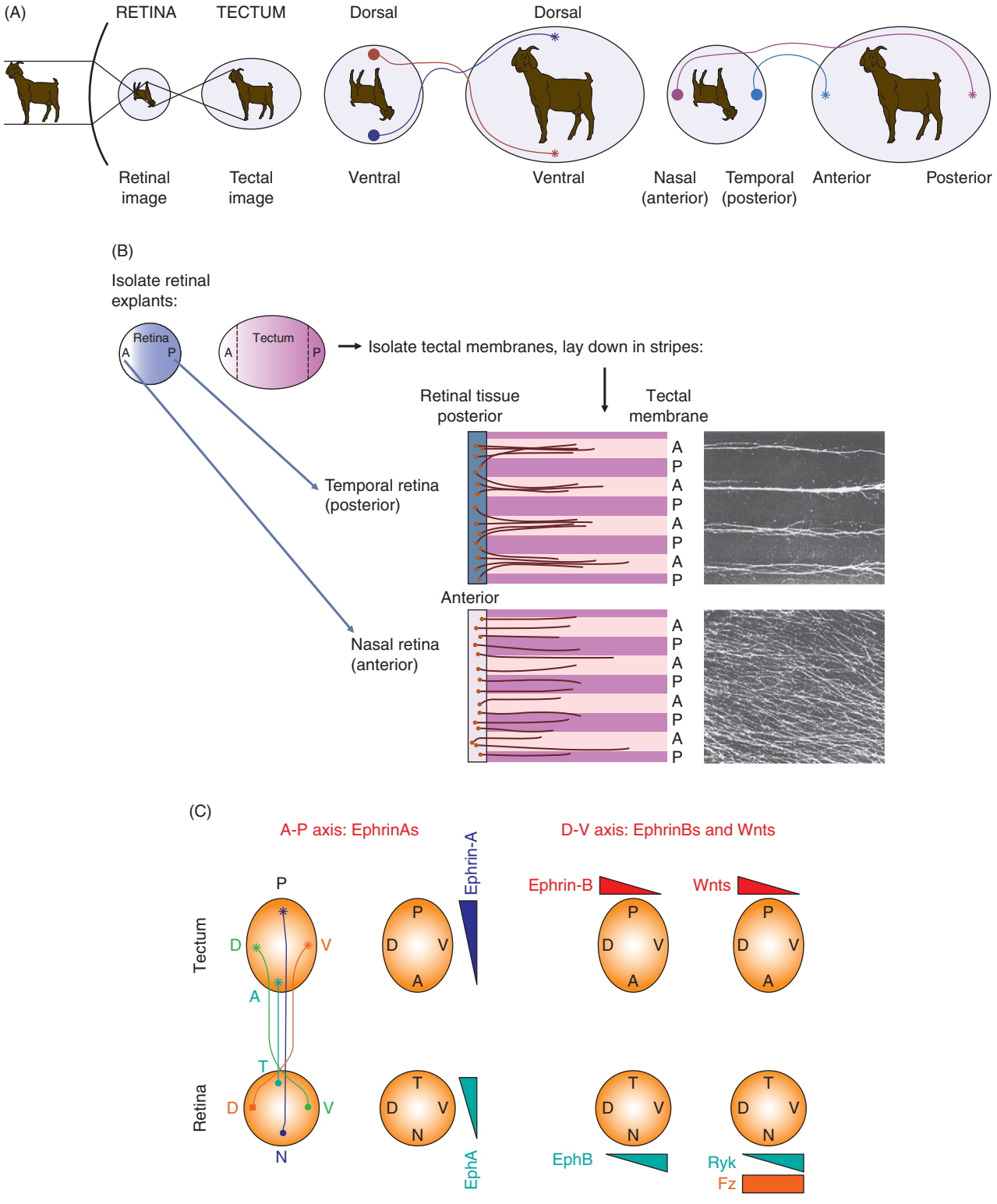
TOPOGRAPHIC MAPPING

So far, we have discussed the discrete trajectories of individual growth cones that navigate a succession of intermediate targets. We now close this chapter with a discussion of the mechanisms that contribute to the development of topographic projections.

In many parts of the nervous system, fields of neurons must project to their targets while maintaining their neighborhood—or "topographic"—relations. A classic example of this behavior is provided by the million or so retinal ganglion cells (RGCs), which are arranged in a two-dimensional sheet within the retina at the back of the eye, and which project to various subcortical targets including the superior colliculus or "optic tectum" in the midbrain. This "retinotectal" projection is topographic, as neighboring neurons in the retina project to neighboring neurons in the target field. It is also inverted along both the anterior-posterior (A-P) and the dorso-ventral (D-V) axes: for example, neurons in the anterior (nasal) retina project to the posterior tectum, while those in the anterior (temporal) retina project to the posterior tectum; similarly dorsal RGCs project to ventral tectum and vice-versa. In this way, the two-dimensional image of the visual world that is projected onto the retina is preserved, albeit inverted, on the tectum (Fig. 16.13A). Topographic projections occur throughout the nervous system and are particularly evident in ascending sensory and descending motor projections.

How do these topographic projections form? In the mid-twentieth century, Roger Sperry (1963) studied the regeneration of the retinotectal projection after optic nerve transection in some amphibians (which, unlike birds and mammals, can regenerate their retinal axons) and found that regenerating axons grow back to their appropriate topographic target regions in the tectum even when the system was perturbed in a variety of ways (e.g., when the eye was rotated in its socket, or part of the retina or tectum was deleted). Impressed with the apparent "hard-wiring" of these projections,

FIGURE 16.13 Topographic mapping of retinal projections in the visual system. (A). The projection of retinal ganglion cells (RGCs) to the optic tectum is topographic. A map of the visual world on the tectum or superior colliculus (left) is inverted with respect to that on the retina because axons from dorsal and ventral retina project to ventral and dorsal tectum, respectively (middle), and axons from anterior (nasal) and posterior (temporal) retina project to posterior and anterior tectum, respectively (right). *Adapted from Tessier-Lavigne (1995).* (B) An *in vitro* assay demonstrates topographically appropriate recognition of tectal membranes by some retinal axons. Small slivers of nasal and temporal retina (top left) are placed on carpets of membranes from tectal cells and allowed to grow. The carpets have alternating stripes of anterior and posterior tectal membranes. In these experiments, the posterior (temporal) retinal axons show an appropriate preference for growth on membranes derived from their correct target (anterior tectum) (Walter et al., 1987b). Other experiments showed that this is due to avoidance of a repellent activity, due to EphrinA ligands, that is enriched in posterior tectum. Note that in this experiment, anterior (nasal) axons do not show the expected preference for posterior tectal membranes, which represents a limitation of the assay; more refined assays are, however, able to demonstrate this. (C) A model for topographic mapping by gradients of guidance molecules and their receptors. Along the A-P axis, retinal neurons have a graded expression of EphA receptors, which enables them to interpret a decreasing posterior-to-anterior gradient of EphrinA ligands on the tectum to select their appropriate target on that axis. Mapping on the D-V axis, in contrast, involves gradients of EphrinB and Wnt ligands that are read by graded distributions of cognate receptors (EphB and Frizzled, respectively; Wnt ligands are also detected by Ryk, which has a nongraded distribution). *Adapted from Flanagan (2006).*



he postulated that a molecular code was likely used to match neurons in each retinal region to the appropriate target region in the tectum. He argued further that it would be difficult to achieve this matching using completely distinct molecular cues for each region, and that instead the guidance information was likely to be present in *gradients* of appropriate recognition molecules. Thus, he conceived that *"an orderly cytochemical mapping in terms of two or more gradients . . . would stamp each cell with its appropriate latitude and longitude expressed in a kind of*

chemical code with matching values between the retinal and tectal map."

In vitro experiments in the 1980s succeeded in demonstrating specificity in the interactions of neurons from particular regions of the retina with appropriate tectal cells, and, in later experiments, even with crude membrane preparations derived from these cells (Walter, Kern-Veits, Huf, Stolze, & Bonhoeffer, 1987a). In a particularly dramatic example, retinal axons were forced to grow on alternating "stripes" of membranes derived from anterior tectal cells and posterior tectal cells (Fig. 16.13B). Remarkably, axons derived from neurons in the temporal (posterior) retina chose to grow on membranes from anterior tectal cells—their appropriate target—rather than posterior tectal membranes (Fig. 16.13B). It was then found that posterior axons choose anterior membranes not because they are specifically attracted to those membranes, but rather because they avoid the posterior membranes, which were shown to contain an activity that is selectively repulsive to temporal axons (Walter, Henke-Fahle, & Bonhoeffer, 1987b). Biochemical analysis showed that the repulsive activity in posterior membranes was mediated by several members of the EphrinA subfamily of axon guidance molecules (introduced earlier in this chapter; Fig. 16.5A), which are present in an increasing anterior to posterior gradient along the tectum—a gradient of repulsion (Fig. 16.13C) (Cheng, Nakamoto, Bergemann, & Flanagan, 1995; Drescher et al., 1995).

These initial observations were eventually generalized to show that topographic mapping along the A-P axis involves recognition of an increasing EphrinA gradient along that axis in the tectum, with neurons at distinct positions along the A-P axis of the retina showing a preference for particular concentrations of EphrinAs by virtue of their graded expression of EphA receptors (Hansen, Dallal, & Flanagan, 2004). Interestingly, topographic mapping along the D-V axis also appears to involve gradients, but of different factors—specifically, countergradients of attractive EphrinBs and repulsive Wnts (Fig. 16.13C). Another interesting twist is that topographic projections can also be generated using other cellular strategies, such as the growth of all axons across a target field followed by the later pruning back of axons in a topographically appropriate manner, with the topographic pruning controlled by similar molecular gradients (e.g., of EphrinAs) (Feldheim & O'Leary, 2010).

Summary

Analysis of the retinotectal system has shown how spatial gradients of axon guidance molecules across a target field can be used to establish topographic projections. The specific cellular strategies can vary, with the recognition molecules either repelling or attracting, or triggering axon branching or pruning, but in all cases the maps are established as a result of neurons responding in a topographically appropriate manner to gradients of guidance cues.

References

Bashaw, G. J., & Klein, R. K. (2010). Signaling from Axon guidance receptors. *Cold Spring Harbor Perspector Biology, 3,* 257–272.

Bastiani, M. J., du Lac, S., & Goodman, C. S. (1886). Guidance of neuronal growth cones in the grasshopper embryo. I. Recognition of a specific axonal pathway by the pCC neuron. *Journal of Neuroscience, 6,* 3518–3531.

Brankatschk, M., & Dickson, B. J. (2006). Netrins guide Drosophila commissural axons at short range. *Nature Neuroscience, 9,* 188–194.

Brose, K., Bland, K. S., Wang, K. H., Arnott, D., Henzel, W., Goodman, C. S., Tessier-Lavigne, M., & Kidd, T. (1999). Slit proteins bind Robo receptors and have an evolutionarily conserved role in repulsive axon guidance. *Cell, 96,* 795–806.

Caudy, M., & Bentley, D. (1986). Pioneer growth cone steering along a series of neuronal and non-neuronal cues of different affinities. *Journal of Neuroscience, 6,* 1781–1795.

Charron, F., Stein, E., Jeong, J., McMahon, A. P., & Tessier-Lavigne, M. (2003). The morphogen sonic hedgehog is an axonal chemoattractant that collaborates with netrin-1 in midline axon guidance. *Cell, 113,* 11–23.

Cheng, H. J., Nakamoto, M., Bergemann, A. D., & Flanagan, J. G. (1995). Complementary gradients in expression and binding of ELF-1 and Mek4 in development of the topographic retinotectal projection map. *Cell, 11,* 371–381.

Chisholm, A., & Tessier-Lavigne, M. (1999). Conservation and divergence of axon guidance mechanisms. *Current Opinion in Neurobiology, 9,* 603–615.

Deiner, M. S., Kennedy, T. E., Fazeli, A., Serafini, T., Tessier-Lavigne, M., & Sretavan, D. W. (1997). Netrin-1 and DCC mediate axon guidance locally at the optic disc: Loss of function leads to optic nerve hypoplasia. *Neuron, 19,* 575–589.

Dent, E. W., Gupton, S. L., & Gertler, F. B. (2010). The growth cone cytoskeleton in axon outgrowth and guidance. *Cold Spring Harbor Perspector Biology, 3,* 199–238.

Dickson, B. J. (2002). Molecular mechanisms of axon guidance. *Science, 298,* 1959–1964.

Drescher, U., Kremoser, C., Handwerker, C., Löschinger, J., Noda, M., & Bonhoeffer, F. (1995). *In vitro* guidance of retinal ganglion cell axons by RAGS, a 25 kDa tectal protein related to ligands from Eph receptor tyrosine kinases. *Cell, 11,* 359–370.

Edelman, G. M. (1986). Cell adhesion molecules in the regulation of animal form and tissue pattern. *Annual Review of Cell and Developmental Biology, 2,* 81–116.

Falk, J., Bechara, A., Fiore, R., Nawabi, H., Zhou, H., Hoyo-Becerra, C., …, Castellani, V. (2005). Dual functional activity of semaphorin 3B is required for positioning the anterior commissure. *Neuron, 48,* 63–75.

Feldheim, D. A., & O'Leary, D. D. (2010). Visual map development: Bidirectional signaling, bifunctional guidance molecules, and competition. *Cold Spring Harbor Perspector Biology, 3,* 115–143.

Flanagan, J. G. (2006). Neural map specification by gradients. *Current Opinion in Neurobiology, 16,* 59–66.

Forscher, P., & Smith, S. J. (1988). Actions of cytochalasins on the organization of actin filaments and microtubules in a neuronal growth cone. *The Journal of Cell Biology, 107,* 1505–1516.

Fuerst, P. G., Koizumi, A., Masland, R. H., & Burgess, R. W. (2008). Neurite arborization and mosaic spacing in the mouse retina require DSCAM. *Nature, 451,* 470–474.

Goodman, C. S., & Shatz, C. J. (1993). Developmental mechanisms that generate precise patterns of neuronal connectivity. *Cell (Cambridge, Mass.), 72,* 77–98.

Hamelin, M., Zhou, M., Su, M. W., Scott, I. M., & Culotti, J. G. (1993). Expression of the UNC-5 guidance receptor in the touch neurons of C. elegans steers their axons dorsally. *Nature, 364,* 327–330.

Hansen, M. J., Dallal, G. E., & Flanagan, J. G. (2004). Retinal axon response to ephrin-As shows a graded, concentration-dependent transition from growth promotion to inhibition. *Neuron, 42,* 717–730.

Hong, K., Hinck, L., Nishiyama, M., Poo, M. M., Tessier-Lavigne, M., & Stein, E. (1999). A ligand-gated association between cytoplasmic domains of UNC5 and DCC family receptors converts netrin-induced growth cone attraction to repulsion. *Cell, 97,* 927–941.

Höpker, V. H., Shewan, D., Tessier-Lavigne, M., Poo, M., & Holt, C. (1999). Growth-cone attraction to netrin-1 is converted to repulsion by laminin-1. *Nature, 401,* 69–73.

Jen, J. C., Chan, W. M., Bosley, T. M., Wan, J., Carr, J. R., Rub, U., …, Engle, E. C. (2004). Mutations in a human ROBO gene disrupt hindbrain axon pathway crossing and morphogenesis. *Science, 304,* 1509–1513.

Kantor, D. B., Chivatakarn, O., Peer, K. L., Oster, S. F., Inatani, M., Hansen, M. J., …, Kolodkin, A. L. (2004). Semaphorin 5A is a bifunctional axon guidance cue regulated by heparan and chondroitin sulfate proteoglycans. *Neuron, 44,* 961–975.

Kennedy, T. E., Serafini, T., de la Torre, J. R., & Tessier-Lavigne, M. (1994). Netrins are diffusible chemotropic factors for commissural axons in the embryonic spinal cord. *Cell, 78,* 425–435.

Kennedy, T. E., Wang, H., Marshall, W., & Tessier-Lavigne, M. (2006). Axon guidance by diffusible chemoattractants: A gradient of netrin protein in the developing spinal cord. *Journal of Neuroscience, 26,* 8866–8874.

Kidd, T., Bland, K. S., & Goodman, C. S. (1999). Slit is the midline repellent for the robo receptor in Drosophila. *Cell, 96,* 785–794.

Kidd, T., Brose, K., Mitchell, K. J., Fetter, R. D., Tessier-Lavigne, M., Goodman, C. S., & Tear, G. (1998). Roundabout controls axon crossing of the CNS midline and defines a novel subfamily of evolutionarily conserved guidance receptors. *Cell, 92,* 205–215.

Klein, R. (2004). Eph/ephrin signaling in morphogenesis, neural development and plasticity. *Current Opinion in Cell Biology, 16,* 580–589.

Kolodkin, A. L., Matthes, D. J., O'Connor, T. P., Patel, N. H., Admon, A., Bentley, D., & Goodman, C. S. (1992). Fasciclin IV: Sequence, expression, and function during growth cone guidance in the grasshopper embryo. *Neuron, 9,* 831–845.

Kolodkin, A. L., & Tessier-Lavigne, M. (2010). Mechanisms and molecules of neuronal wiring: A primer. Cold Spring Harbor Perspector *Biology, 3,* 1–14.

Kruger, R. P., Aurandt, J., & Guan, K. L. (2005). Semaphorins command cells to move. *Nature Reviews. Molecular Cell Biology, 6,* 789–800.

Lalli, G., & Hall, A. (2010). Rho and Ras GTPases in axon growth, guidance, and branching. *Cold Spring Harbor Perspectives in Biology, 3,* 239–56.

Lance-Jones, C., & Landmesser, L. (1980). Motoneurone projection patterns in the chick hind limb following early partial reversals of the spinal cord. *The Journal of Physiology (London), 302,* 581–602.

Li, H. S., Chen, J. H., Wu, W., Fagaly, T., Zhou, L., Yuan, W., …, Rao, Y. (1999). Vertebrate slit, a secreted ligand for the transmembrane protein roundabout, is a repellent for olfactory bulb axons. *Cell, 96,* 807–818.

Lin, C. H., Espreafico, E. M., Mooseker, M. S., & Forscher, P. (1996). Myosin drives retrograde F-actin flow in neuronal growth cones. *Neuron, 16,* 769–782.

Lin, D. M., Fetter, R. D., Kopczynski, C., Grenningloh, G., & Goodman, C. S. (1994). Genetic analysis of fasciclin II in Drosophila: Defasciculation, refasiculation, and altered fasciculation. *Neuron, 13,* 1055–1069.

Lin, C. -H., Thompson, C. A., & Forscher, P. (1994). Cytoskeletal reorganization underlying growth cone motility. *Current Opinion in Neurobiology, 4,* 640–647.

Liu, Y., Lu, C. C., Wang, Z. B., Lyuksyutova, A. I., Song, X. J., & Zou, Y. (2005). Ryk-mediated Wnt repulsion regulates posterior-directed growth of corticospinal tract. *Nature Neuroscience, 8,* 1151–1159.

Luo, Y., Raible, D., & Raper, J. A. (1993). Collapsin: A protein in brain that induces the collapse and paralysis of neuronal growth cones. *Cell, 75,* 217–227.

Lyuksyutova, A. I., Lu, C. C., Milanesio, N., King, L. A., Guo, N., Wang, Y., …, Zou, Y. (2003). Anterior-posterior guidance of commissural axons by Wnt-frizzled signaling. *Science, 302,* 1984–1988.

Ma, L., & Tessier-Lavigne, M. (2007). Dual branch-promoting and branch-repelling actions of Slit/Robo signaling on peripheral and central branches of developing sensory axons. *Journal of Neuroscience, 27,* 6843–6851.

Matise, M. P., Lustig, M., Sakurai, T., Grumet, M., & Joyner, A. L. (1999). Ventral midline cells are required for the local control of commissural axon guidance in the mouse spinal cord. *Development, 126,* 3649–3659.

McConnell, S. K., Ghosh, A., & Shatz, C. J. (1994). Subplate pioneers and the formation of descending connections from the cerebral cortex. *Journal of Neuroscience, 14,* 1892–1907.

Ming, G. -L., Song, H. -J., Berninger, S., Holt, C. E., Tessier-Lavigne, M., & Poo, M. -M. (1997). cAMP-dependent growth cone guidance by netrin-1. *Neuron, 19,* 1225–1235.

Pasterkamp, R. J., & Giger, R. J. (2009). Semaphorin function in neural plasticity and disease. *Current Opinion in Neurobiology, 19,* 263–274.

Polleux, F., Giger, R. J., Ginty, D. D., Kolodkin, A. L., & Ghosh, A. (1998). Patterning of cortical efferent projections by semaphorin-neuropilin interactions. *Science, 282,* 1904–1906.

Polleux, F., Morrow, T., & Ghosh, A. (2000). Semaphorin 3A is a chemoattractant for cortical apical dendrites. *Nature, 404,* 567–573.

Ramón y Cajal, S. (1890). Sur l'origine et les ramifications des fibres nerveuses de la moelle embryonaire. *Anatomischer Anzeiger, 5,* 609–613. Extract from Ramon y Cajal, S. (1909). *Histology of the nervous system* (N. Swanson & L. W. Swanson, Trans.). Oxford: Oxford Univ. Press, 1995.

Raper, J. A. (2000). Semaphorins and their receptors in vertebrates and invertebrates. *Opinion in Neurobiology, 10,* 88–94.

Raper, J. A., Bastiani, M. J., & Goodman, C. S. (1984). Pathfinding by neuronal growth cones in grasshopper embryos. IV. The effects of ablating the A and P axons upon the behavior of the G growth cone. *Journal of Neuroscience, 4,* 2329–2345.

Sabatier, C., Plump, A. S., Le, M., Brose, K., Tamada, A., Murakami, F., …, Tessier-Lavigne, M. (2004). The divergent Robo family protein rig-1/Robo3 is a negative regulator of slit responsiveness required for midline crossing by commissural axons. *Cell, 117,* 157–169.

Schmitt, A. M., Shi, J., Wolf, A. M., Lu, C. C., King, L. A., & Zou, Y. (2006). Wnt-Ryk signalling mediates medial-lateral retinotectal topographic mapping. *Nature, 439,* 31–37.

Schmucker, D., Clemens, J. C., Shu, H., Worby, C. A., Xiao, J., Muda, M., …, Zipursky, S. L. (2000). Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell, 101,* 671–684.

Serafini, T., Colamarino, S. A., Leonardo, E. D., Wang, H., Beddington, R., Skarnes, W., & Tessier-Lavigne, M. (1996). Netrin-1 is required for commissural axon guidance in the developing vertebrate nervous system. *Cell, 87,* 1001–1014.

Serafini, T., Kennedy, T. E., Galko, M. J., Mirzayan, C., Jessell, T. M., & Tessier-Lavigne, M. (1994). The netrins define a family of axon outgrowth-promoting proteins homologous to C. elegans UNC-6. *Cell (Cambridge, Mass.), 78,* 409–424.

Song, H., Ming, G., He, Z., Lehmann, M., McKerracher, L., Tessier-Lavigne, M., & Poo, M. (1998). Conversion of neuronal growth cone responses from repulsion to attraction by cyclic nucleotides. *Science, 281,* 1515–1518.

Sperry, R. W. (1963). Chemoaffinity in the orderly growth of nerve fiber patterns and connections. *Proceedings of the National Academy of Sciences of the United States of America, 50,* 703–710.

Srour, M., Rivière, J. B., Pham, J. M., Dubé, M. P., Girard, S., Dion, P. A., …, Rouleau, G. A. (2010). Mutations in DCC cause congénital Mirror movements. *Science, 328,* 592.

Stein, E., & Tessier-Lavigne, M. (2001). Hierarchical organization of guidance receptors: Slit silences netrin attraction through a Robo/DCC receptor complex. *Science, 291,* 1847–2034.

Suter, D. M., & Forscher, P. (2000). Substrate-cytoskeletal coupling as a mechanism for the regulation of growth cone motility and guidance. *Journal of Neurobiology, 44,* 97–113.

Tamagnone, L., & Comoglio, P. M. (2000). Signaling by semaphorin receptors: Cell guidance and beyond. *Trends in Cell Biology, 10,* 377–383.

Taniguchi, M., Yuasa, S., Fujisawa, H., Naruse, I., Saga, S., Mishina, M., & Yagi, T. (1997). Disruption of semaphorin III/D gene causes severe abnormality in peripheral nerve projection. *Neuron, 19,* 519–530.

Tessier-Lavigne, M. (1995). Eph receptor tyrosine kinases, axon repulsion, and the development of topographic maps. *Cell, 11,* 345–348.

Tessier-Lavigne, M., & Goodman, C. S. (1996). The molecular biology of axon guidance. *Science, 274,* 1123–1133.

Tessier-Lavigne, M., Placzek, M., Lumsden, A. G., Dodd, J., & Jessell, T. M. (1988). Chemotropic guidance of developing axons in the mammalian central nervous system. *Nature (London), 336,* 775–778.

Tran, T. S., Kolodkin, A. L., & Bharadwaj, R. (2007). Semaphorin regulation of cellular morphology. *Annual Review of Cell and Developmental Biology, 23,* 263–292.

Van Vactor, D., Wall, D. P., & Johnson, K. G. (2006). Heparan sulfate proteoglycans and the emergence of neuronal connectivity. *Current Opinion in Neurobiology, 16,* 40–51.

Wadsworth, W. G., Bhatt, H., & Hedgecock, E. M. (1996). Neuroglia and pioneer neurons express UNC-6 to provide global and local netrin cues for guiding migrations in C. elegans. *Neuron, 16,* 35–46.

Walter, J., Henke-Fahle, S., & Bonhoeffer, F. (1987b). Avoidance of posterior tectal cell membranes by temporal retinal axons. *Development, 101,* 909–913.

Walter, J., Kern-Veits, B., Huf, J., Stolze, B., & Bonhoeffer, F. (1987a). Recognition of position-specific properties of tectal cell membranes by retinal axons *in vitro. Development, 101,* 685–696.

Wang, K. H., Brose, K., Arnott, D., Kidd, T., Goodman, C. S., Henzel, W., & Tessier-Lavigne, M. (1999). Biochemical purification of a mammalian slit protein as a positive regulator of sensory axon elongation and branching. *Cell, 96,* 771–784.

Wessells, N. K., & Nuttall, R. P. (1978). Normal branching, induced branching, and steering of cultured parasympathetic motor neurons. *Experimental Cell Research, 115,* 111–122.

Whitford, K. L., Dijkhuizen, P., Polleux, F., & Ghosh, A. (2002). Molecular control of cortical dendrite development. *Annual Review of Neuroscience, 25,* 127–149.

Winberg, M. L., Noordermeer, J. N., Tamagnone, L., Comoglio, P. M., Spriggs, M. K., Tessier-Lavigne, M., & Goodman, C. S. (1998). Plexin A is a neuronal semaphorin receptor that controls axon guidance. *Cell, 95,* 903–916.

Zhan, X. L., Clemens, J. C., Neves, G., Hattori, D., Flanagan, J. J., Hummel, T., …, Zipursky, S. L. (2004). Analysis of Dscam diversity in regulating axon guidance in Drosophila mushroom bodies. *Neuron, 43,* 673–686.

Zipursky, S. L., & Sanes, J. R. (2010). Chemoaffinity revisited: Dscams, protocadherins, and neural circuit assembly. *Cell, 143,* 343–353.

Zou, Y., Stoeckli, E., Chen, H., & Tessier-Lavigne, M. (2000). Squeezing axons out of the gray matter: A role for Slit and Semaphorin proteins from midline and ventral spinal cord. *Cell, 102,* 363–365.

Alex L. Kolodkin and Marc Tessier-Lavigne

CHAPTER

17

Synapse Formation

A fundamental issue in neurobiology is defining the mechanisms by which neurons form synapses with their targets. The development of appropriate synaptic connections requires a series of steps including the specification and generation of neurons and their target cells (Chapters 13, 14, and 15), the guidance of axons to their targets, the selection of appropriate targets, the formation of orderly and specific projections within the target (Chapter 16), and ultimately the induction of a specialized presynaptic terminal and postsynaptic membrane (this chapter). The first section of this chapter focuses on neuromuscular synapses that form between motor neurons and muscle cells in the peripheral nervous system. The second section summarizes current views of how neuron–neuron synapses in the central nervous system are assembled.

DEVELOPMENT OF THE NEUROMUSCULAR SYNAPSE

Once axons have arrived at their appropriate target destination, synapse formation ensues. Much of our understanding about the mechanisms of synapse formation arises from studies of the neuromuscular synapse. These studies have benefited from (1) the relative ease of experimentally manipulating developing and regenerating neuromuscular synapses *in vivo*, (2) cell culture systems for both motor neurons and skeletal muscle cells, (3) the *Torpedo* electric organ, an abundant and homogeneous source of neuromuscular-like synapses (Box 17.1), and (4) transgenic and mutant mice for studying and altering gene expression. Consequently, we have a good, although incomplete, understanding of the mechanisms that lead to the formation of the neuromuscular synapse.

The assembly of neuromuscular synapses is a multistep process. It requires coordinated interactions between motor neurons and muscle fibers, which eventually lead to the formation of a highly specialized postsynaptic membrane and a highly differentiated nerve terminal. As a consequence, acetylcholine receptors (AChRs) become highly concentrated in the postsynaptic membrane and arranged in perfect register with active zones in the presynaptic nerve terminal, ensuring fast, robust, and reliable synaptic transmission. Because neuromuscular synapses control all voluntary movement as well as respiration, the formation and function of this synapse are essential for survival. Disruption of this synapse, caused by axon withdrawal and motor neuron death, is responsible for amyotrophic lateral sclerosis (ALS). A large number of proteins contribute to the process of synapse formation, several of which play fundamental roles: (1) Agrin, a motor neuron-derived ligand essential to stabilize nascent synapses; (2) Lrp4, the receptor for Agrin; (3) MuSK, the receptor tyrosine kinase that transduces the Agrin signal; (4) Dok-7, an adaptor protein that is recruited to activated MuSK; and (5) Rapsyn, the AChR-associated protein that anchors AChRs at synapses. Defects in this signaling pathway, which lead to a reduced number of AChRs at synapses, are responsible for a variety of congenital neuromuscular disorders, termed congenital myasthenia, and autoantibodies to AChRs, Lrp4 or MuSK are responsible for myasthenia gravis. Here, we describe the structure of the neuromuscular synapse and the molecules and mechanisms responsible to assemble this structure.

Substructural Organization of the Neuromuscular Synapse

An adult myofiber, a syncitial cell containing hundreds to thousands of nuclei, is innervated by a single

© 2013 Elsevier Inc. All rights reserved.

BOX 17.1

TORPEDO ELECTRIC ORGAN

The majority of proteins known to be localized to neuromuscular synapses were first identified in postsynaptic membranes isolated from the electric organ of the marine ray *Torpedo*. Indeed, this specialized tissue has been essential for the identification and purification of AChR, AChE, Rapsyn, Syntrophin, Agrin, MuSK, and several synaptic vesicle proteins. The electric organ is a particularly homogeneous and abundant source of pre- and postsynaptic membranes that are similar in structure and function to those at neuromuscular synapses. The biochemical advantages of the electric organ are evident from its anatomy. The postsynaptic cell, which differentiates initially as a syncitial skeletal muscle fiber but subsequently loses its contractile machinery, is termed an electroplaque. Each electroplaque is a thin, elongated cell (1 cm × 1 cm × 10 μm) that is so densely innervated that nearly one-half of the electrocyte membrane is studded by nerve terminals. In contrast, nerve terminals occupy less than 0.1% of the cell surface of a skeletal myofiber. Because of this dense innervation and because the electric organ from a moderate size ray weighs several kilograms, it is possible to obtain several mg of purified postsynaptic proteins from a single electric organ.

Innervation is restricted to the ventral surface of the electroplaque, whereas the dorsal surface is enriched for the sodium/potassium ATPase that maintains the resting potential. Because the electroplaque lacks action potentials, activation of AChRs on the innervated ventral surface results in a voltage drop across the innervated but not the noninnervated membrane of the electrocyte. Since thousands of electroplaques are stacked closely one upon another, the potential difference across a single electrocyte is summated by the stack of electroplaques, resulting in a several thousand volt potential difference across the entire electric organ, a voltage that is sufficient to stun prey.

Steven J. Burden and Peter Scheiffele

motor axon that terminates and arborizes over approximately 0.1% of the muscle fiber's cell surface. The neurotransmitter receptor, AChR, is highly concentrated in this small patch of the muscle fiber membrane, and its precise localization to synaptic sites during development is a hallmark of the inductive events of synapse formation. Although other proteins are likewise concentrated at synaptic sites, much of our knowledge about synaptic differentiation has come from studies designed to understand how AChRs accumulate at synaptic sites.

Nerve terminals are situated in shallow depressions of the muscle cell membrane, which is invaginated further into deep and regular folds, termed postjunctional folds (Fig. 17.1). AChRs and additional proteins (see later) are localized to the crests of these postjunctional folds, whereas other proteins, including sodium channels, are enriched in the troughs of the postjunctional folds. The nerve terminal likewise is organized spatially, and its substructural organization reflects that of the postsynaptic membrane. Synaptic vesicles are sparse in the region of the nerve terminal underlying Schwann cells, which cover nerve terminals and myelinate motor axons, and abundant in the region of the nerve terminal facing the muscle fiber. Moreover, synaptic vesicles are clustered adjacent to a poorly characterized specialization of the presynaptic membrane, termed active zones, which are the sites of synaptic vesicle fusion. Active zones are organized at regular intervals and are aligned precisely with the mouths of the postjunctional folds. This precise registration of active zones and postjunctional folds ensures that acetylcholine encounters a high concentration of AChRs within microseconds after release, thereby facilitating synaptic transmission. The alignment of structural specializations in pre- and postsynaptic membranes, separated by a 500-Å synaptic cleft, suggests that spatially restricted signaling between pre- and postsynaptic cells is important to coordinate pre- and postsynaptic differentiation. The mechanisms that align active zones and postjunctional folds are not well understood, but $\alpha 3/\beta 2$ laminin, produced by the muscle and found in the synaptic basal lamina, binds calcium channels in nerve terminals, and this interaction has a role in anchoring active zones and coordinating pre- and postsynaptic differentiation (Nishimune, Sanes, & Carlson, 2004).

The precise organization of molecules in pre- and postsynaptic membranes belies the concept that the neuromuscular synapse is a simple synapse. Rather, the substructure of pre- and postsynaptic membranes, together with the faithful registration of pre- and postsynaptic specializations, suggests that complex mechanisms are required to assemble the synapse and to coordinate pre- and postsynaptic differentiation.

Muscle-Autonomous Patterning of AChR Expression

As motor axons extend toward developing muscle, muscle cells are themselves undergoing differentiation.

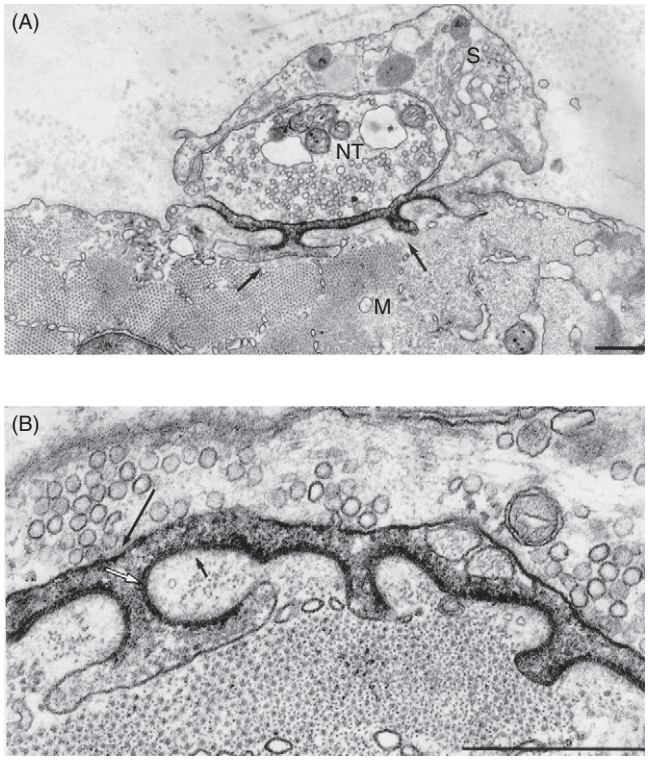


FIGURE 17.1 **Pre- and postsynaptic membranes at the neuromuscular synapse are highly specialized.** (A) An electron micrograph of a neuromuscular synapse shows that the nerve terminal is capped by a Schwann cell and is situated in a shallow depression of the muscle cell membrane, which is invaginated further into deep and regular folds, termed postjunctional folds (arrows). AChRs, labeled with α-bungarotoxin coupled to horseradish peroxidase, are concentrated at the synaptic site. (B) A higher magnification view shows that AChRs are concentrated at the crests and along the sides of the postjunctional folds (white arrow). Rapsyn, Neuregulin receptors, and MuSK are also concentrated in the postsynaptic membrane, whereas Agrin, Neuregulin-1, acetylcholinesterase, S-laminin, and certain isoforms of collagen are localized to the synaptic basal lamina. The postjunctional folds of the myofiber are spaced at regular intervals and are situated directly across from active zones and clusters of synaptic vesicles in the nerve terminal (arrows).

Myotubes, formed by fusion of precursor myoblasts, develop prior to innervation and continue to grow after innervation by further myoblast fusion. Because myoblasts fuse to developing myotubes at their growing ends, the central region of the muscle is more mature than the distal ends of the muscle. Recent studies have shown that the central region of muscle is regionally specialized prior to innervation. For example, AChRs are already clustered in the central region of the muscle prior to and independent of innervation (Flanagan-Steet, Fox, Meyer, & Sanes, 2005; Lin et al., 2001; Panzer, Song, Balice-Gordon, 2006; Yang et al., 2001; Yang, Li, Prescott, Burden, & Wang, 2000;). As such, motor axons approach muscles that are regionally specialized, or prepatterned, prior to innervation, in a manner that preconfigures the prospective zone of innervation (Fig. 17.2).

How the prospective synaptic region of the muscle becomes prepatterned is not well understood (see following), but muscle prepatterning restricts motor axon growth and promotes synapse formation in the central region of mammalian muscle (Kim & Burden, 2008).

The Synaptic Basal Lamina Contains Signals for Synaptic Differentiation

Following contact with the growth cone of a developing motor neuron, developing muscle fibers undergo a complex differentiation program in the synaptic region, which is dependent upon motor neuron–derived signals. In addition, signals from the muscle regulate differentiation of presynaptic nerve terminals. The idea that signals for regulating both presynaptic and postsynaptic differentiation are contained in the synaptic basal lamina arose from studies of regenerating neuromuscular synapses. Following damage to a motor axon, the distal portion of the motor axon degenerates, and the proximal end regenerates to muscle. The regenerated axon precisely reinnervates the original synaptic site and forms a synapse that is indistinguishable from the original synapse. Original synaptic sites are not thought to provide guidance cues to motor axons; rather, it is believed that the vacated perineural tubes, containing Schwann cells and their basal lamina, provide a favorable substrate for motor axons and have a role in directing regenerating motor axons to original synaptic sites (Son & Thompson, 1995). Although motor axons precisely reinnervate original synaptic sites, there appears to be little, if any, selectivity among motor neurons or among Schwann cells in ensuring accurate regeneration; indeed, both original and foreign motor neurons can accurately and functionally reinnervate original synaptic sites in denervated muscle.

Following damage to motor axons and muscle, nerve terminals and muscle fibers degenerate and are phagocytized, but the basal lamina of the muscle fiber remains intact. Even in the absence of nerve terminals and muscle fibers, several structures, including the terminal Schwann cells, the basal lamina of the postjunctional folds, and AChE, remain at the original synaptic site and allow for its identification.

Axons eventually regenerate into the muscle, and new myofibers regenerate within the basal lamina of the original myofiber. The regenerated motor axons form synapses with the regenerated myofibers precisely at the original synaptic sites. If axons regenerate into muscle, but muscle regeneration is prevented, axons still unerringly reinnervate the original synaptic site on the basal lamina, and active zones form in register with the basal lamina of the original postjunctional folds (Fig. 17.3) (Sanes, Marshall, & McMahan, 1978). Thus, the

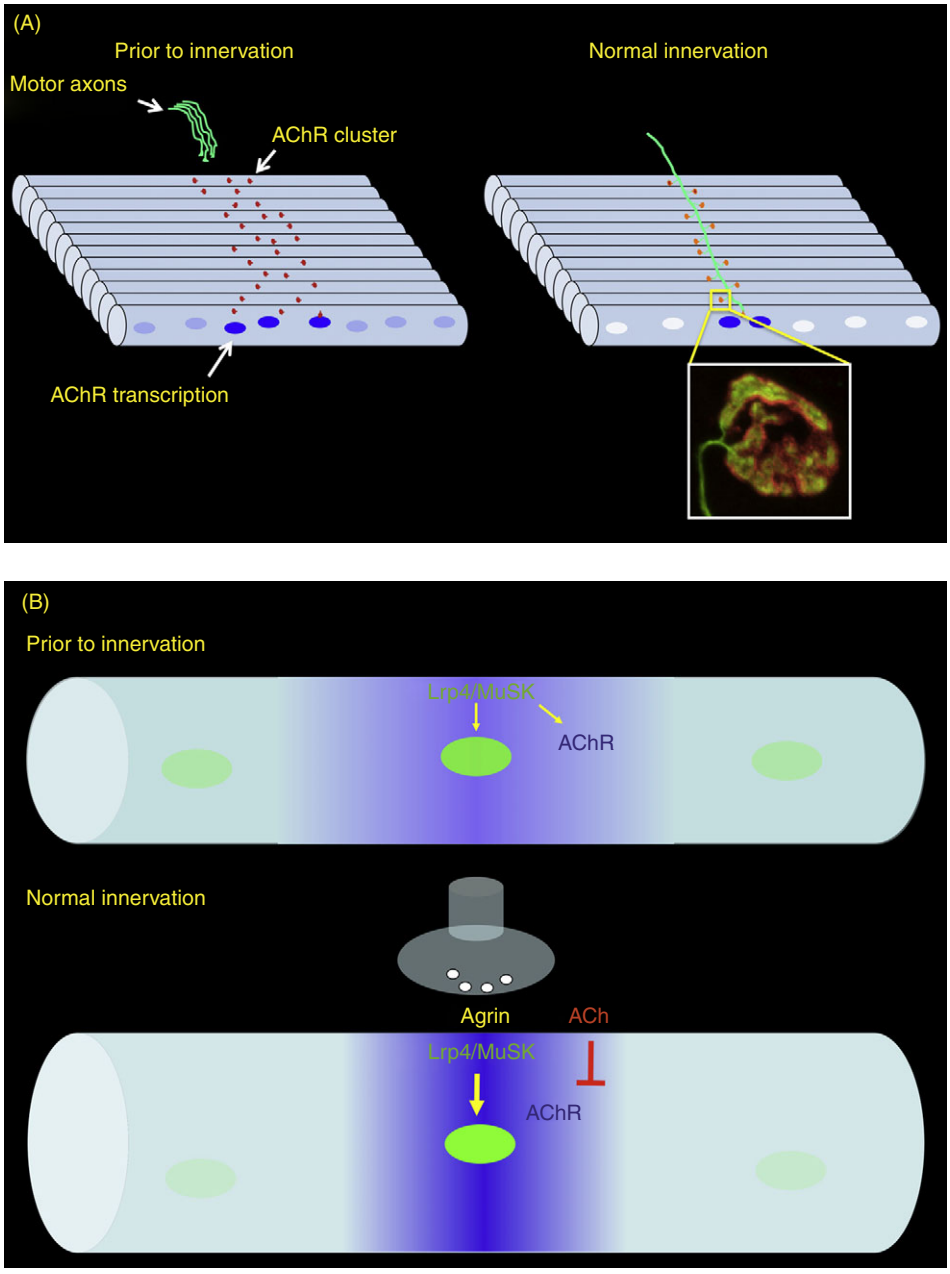


FIGURE 17.2 **Muscle prepatterning.** (A) Motor axons (green) approach muscles that are prepatterned, as *AChR* transcription (blue) and AChR clustering (red) are enhanced in the central, prospective synaptic region of muscle, prior to and independent of innervation. Motor axons terminate and form synapses (orange) in the prepatterned region. (B) Lrp4 and MuSK are required for prepatterning AChRs and are themselves expressed in the central region of muscle independent of innervation. The prepatterned zone of *AChR* expression becomes refined and sharpened following innervation, so that *AChR* expression and AChR clustering are restricted to nascent synaptic sites. Two neural signals refine AChR expression: Agrin/MuSK signaling stabilizes and stimulates AChR clustering, whereas acetylcholine (ACh) disperses AChR clusters. As a consequence of these two signals, AChR clusters are maintained only at nascent synaptic sites.

presence of the myofiber is necessary neither for precise reinnervation nor for the morphological differentiation of regenerated nerve terminals. The vacated perineurial tubes, containing Schwann cells and their basal lamina, may direct regenerating motor axons to original synaptic sites, but the induction of active zones suggests that cues in the synaptic basal lamina, possibly $\alpha 3/\beta 2$ laminin, have a role in organizing the presynaptic terminal.

If regeneration of motor axons is prevented, but myofibers are allowed to regenerate, AChRs accumulate and membrane folds form in the regenerated myofiber precisely at the original synaptic site on the basal lamina (Burden, Sargent, & McMahan, 1979). Thus, in the absence of nerve terminals, myofibers, and terminal Schwann cells, information remains at the original synaptic site and instructs both pre- and postsynaptic differentiation (Fig. 17.3). Because the synaptic basal lamina is the most prominent extracellular structure remaining at neuromuscular synapses following removal of all cells, these results indicate that the synaptic basal lamina contains signals that can induce differentiation of both nerve terminals and myofibers.

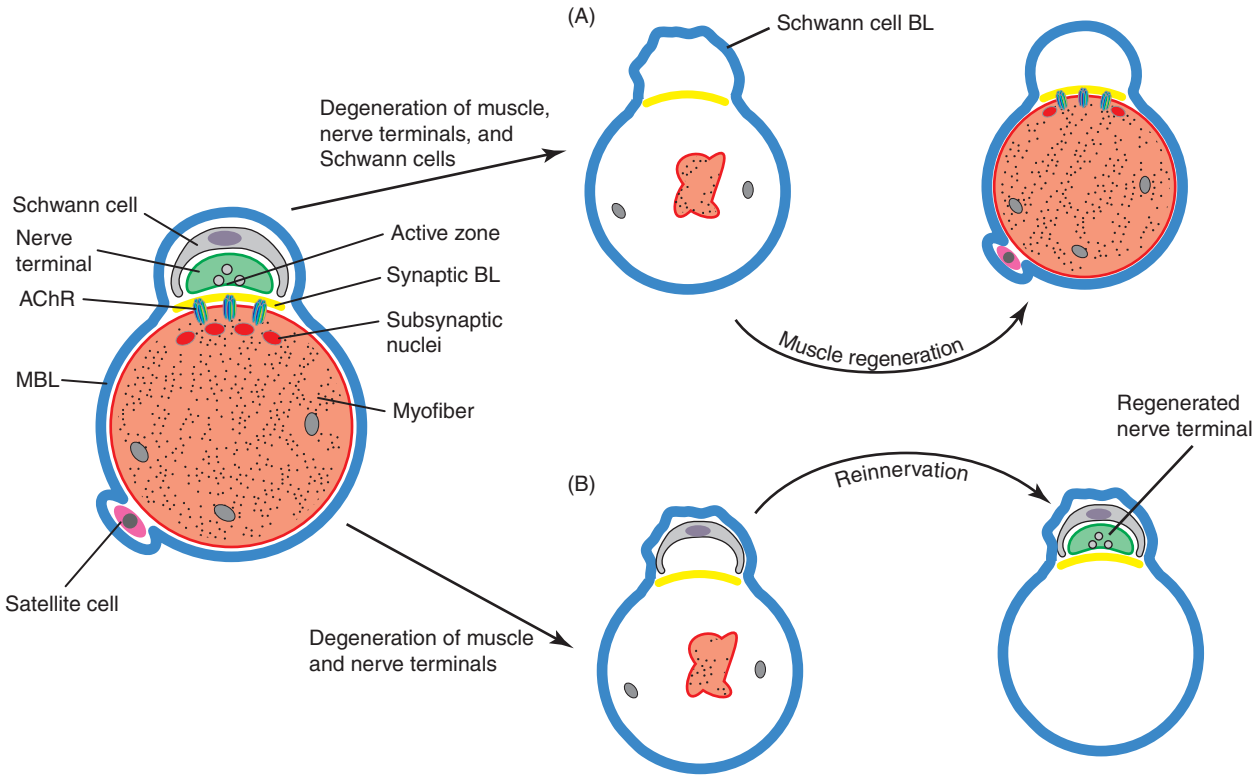


FIGURE 17.3 **The synaptic basal lamina contains signals for presynaptic and postsynaptic differentiation.** A cross section of normal muscle containing a nerve terminal, Schwann cell, and myofiber is shown at the far left. *AChR* genes are induced in synaptic nuclei (red) and AChRs (green) are concentrated at synaptic sites. Active zones and clusters of synaptic vesicles in the nerve terminal are aligned with postjunctional folds in the myofiber. Each myofiber is ensheathed by a basal lamina (MBL), which is specialized at the synaptic site (synaptic BL) and which remains intact following degeneration of the original myofiber. The MBL serves as a scaffold for regenerating myofibers, which form from the fusion of satellite cells that proliferate following damage to the original myofiber. (A) Experiments showing that the synaptic basal lamina contains signals for clustering AChRs and activating *AChR* genes in synaptic nuclei. Following damage and degeneration of axons, Schwann cells, and myofibers, new myofibers regenerate within the basal lamina of the original myofiber in the absence of the nerve. AChRs cluster and *AChR* gene expression is reinduced at original synaptic sites in myofibers that regenerate in the absence of the nerve and other original presynaptic cells. (B) Experiments showing that the synaptic basal lamina contains signals for inducing presynaptic differentiation. Following damage and degeneration of axons and myofibers, motor axons regenerate to original synaptic sites in the absence of the myofiber and accumulate synaptic vesicles precisely across from the sites of the original postjunctional folds.

Agrin is a Signal for Postsynaptic Differentiation

Because clustering of AChRs, unlike the formation of active zones, can be studied readily in cell culture, it has been far simpler to identify the basal lamina signals that induce postsynaptic rather than presynaptic differentiation. Extracellular matrix from the *Torpedo* electric organ, a tissue that is homologous to muscle but more densely innervated, contains an activity that stimulates AChR clustering in cultured myotubes (see Box 17.1). McMahan and colleagues purified the electric organ activity, which they termed Agrin, and showed that Agrin is synthesized by motor neurons, transported in motor axons to synaptic sites, and deposited in the synaptic basal lamina. Agrin also stimulates the clustering of other synaptic proteins, including AChE, Rapsyn, and Utrophin (see later), indicating that Agrin has a central role in synaptic differentiation.

Agrin is a ~200-kDa protein containing multiple epidermal growth factor (EGF)-like signaling domains, two different laminin-like domains, and multiple follistatin-like repeats. The four EGF-like domains and three laminin G domains are contained in the carboxyl-terminal region, which is sufficient for inducing AChR clusters in cultured myotubes; sequences in the amino-terminal region are responsible for the association of Agrin with the extracellular matrix.

The *agrin* gene is expressed in a variety of cell types. Alternative splicing results in multiple Agrin isoforms that differ in their AChR clustering efficiency. The isoform that is most active in clustering AChRs is expressed in neurons, including motor neurons, whereas other Agrin isoforms are expressed in additional cell types, including skeletal muscle cells. The active, neuronal-specific isoforms of Agrin contain 8, 11, or 19 amino acids at a splice site, referred to as the Z site in rat Agrin and the B site in chick Agrin. The *agrin*

gene contains multiple promoters. Although the promoter used in motor neurons leads to expression of isoforms that are associated with the extracellular matrix, the promoter used in the CNS leads to expression of membrane-bound isoforms (Burgess, Skarnes, & Sanes, 2000).

In mice lacking Agrin, AChRs are initially prepatterned, and synapses appear to form, but only transiently (Lin et al., 2001; Misgeld, Kummer, Lichtman, & Sanes, 2005). Over the course of several days, the AChR prepattern is dispersed, synapses are lost, and motor axons begin to grow throughout the muscle. Thus, Agrin is necessary to stabilize the AChR prepattern and nascent synapses. Because the AChR prepattern is stable in the absence of innervation but dispersed by motor axons lacking Agrin, neural signals, other than Agrin, extinguish the AChR prepattern (Lin et al., 2001; Yang et al., 2001). Genetic evidence indicates that acetylcholine is a key neural signal that disperses AChR clusters (Lin et al., 2005; Misgeld et al., 2005). Thus, during normal development, Agrin counteracts acetylcholine-mediated dispersion, thereby stabilizing AChR clusters selectively at nascent synapses (see following).

MuSK and Lrp4 Are Required for Agrin-Mediated Signaling and Synapse Formation

The mechanisms of Agrin-mediated AChR clustering are poorly understood, but a receptor tyrosine kinase, termed MuSK, is a critical component of an Agrin receptor complex. MuSK is expressed in *Torpedo* electric organ and in skeletal muscle, where it is concentrated in the postsynaptic membrane (Jennings, Dyer, & Burden, 1993; Valenzuela et al., 1995). Mice deficient in MuSK (DeChiara et al., 1996), like *agrin* mutant mice, lack normal neuromuscular synapses. Both *agrin* and *MuSK* mutant mice are immobile, cannot breathe, and die at birth. Muscle differentiation is normal in *agrin* and *MuSK* mutant mice, but muscle fibers in *MuSK* mutant mice lack all known features of postsynaptic differentiation. Moreover, unlike *agrin* mutant mice, the AChR prepattern is absent from *MuSK* mutant mice at all stages, indicating that MuSK is essential not only for synapse formation but also to establish muscle prepatterning. Muscle-derived proteins, including AChRs and AChE, which are concentrated at synapses in normal mice, are distributed uniformly in *MuSK* mutant myofibers (Fig. 17.4). In addition, *AChR* genes, which are transcribed selectively in synaptic nuclei of normal muscle fibers (see later), are transcribed at similar rates in synaptic and nonsynaptic nuclei of muscle fibers from *MuSK* mutant mice.

Multiple lines of evidence indicate that MuSK is required for Agrin-mediated signaling and is a component of an Agrin receptor complex (Glass et al., 1996): (1) Agrin can be chemically cross-linked to MuSK in cultured myotubes; (2) Agrin induces rapid tyrosine phosphorylation

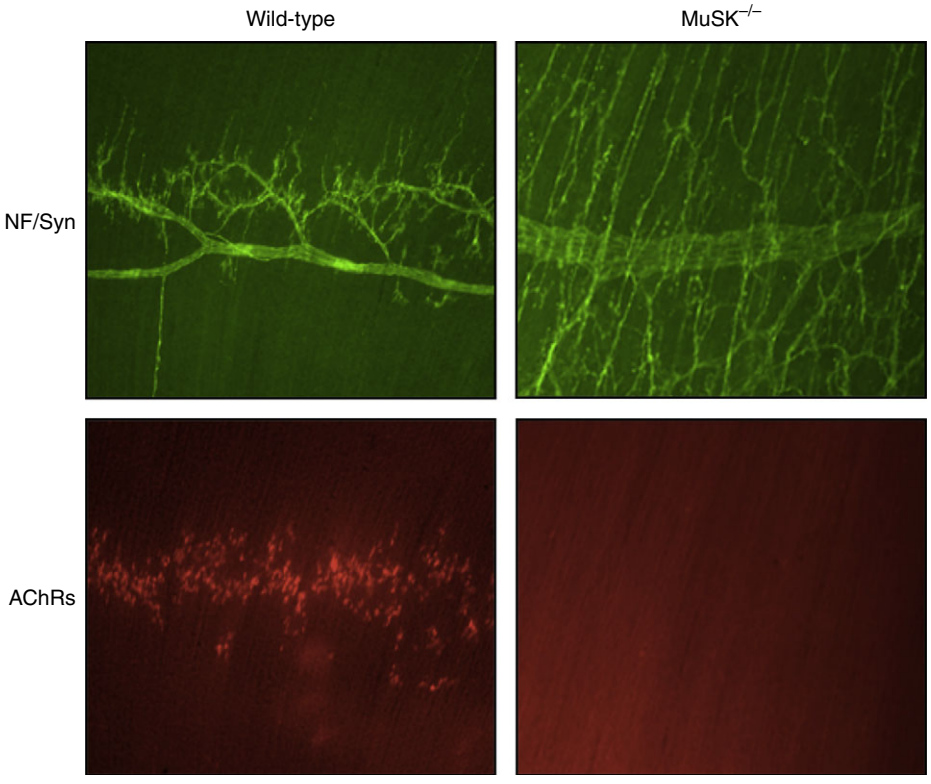


FIGURE 17.4 **Presynaptic and postsynaptic differentiation are defective in mice lacking MuSK.** Whole mounts of muscle from wild-type and *MuSK* mutant mice were stained with antibodies to neurofilament (NF) and synaptophysin (Syn) to label motor axons and nerve terminals (green), respectively, and with α-bungarotoxin to label acetylcholine receptors (AChRs) (red). In *MuSK* mutant mice, motor axons fail to stop and differentiate adjacent to the main intramuscular nerve and instead wander aimlessly over the muscle; AChRs, as well as other postsynaptic proteins, are expressed at normal levels but they fail to cluster.

of MuSK in cultured myotubes; (3) a recombinant, soluble extracellular fragment of MuSK inhibits Agrin-induced AChR clustering in cultured muscle cells; (4) cultured *MuSK* mutant muscle cells, unlike normal muscle cells, do not cluster AChRs in response to Agrin; and (5) dominant-negative forms of MuSK inhibit Agrin-induced AChR clustering in cultured myotubes. MuSK itself, however, does not bind Agrin (Glass et al., 1996).

Instead, the muscle receptor for Agrin is the low-density lipoprotein receptor-related protein 4 (Lrp4) (Kim et al., 2008; Zhang et al., 2008). Neural Agrin binds Lrp4, which stimulates association between Lrp4 and MuSK and increases MuSK kinase activity (Kim et al., 2008; Zhang et al., 2008). Only neural forms of Agrin bind Lrp4, providing an explanation for the selective activation of MuSK by neural isoforms of Agrin (Kim et al., 2008). Because Lrp4 self associates, independent of Agrin, Agrin may increase MuSK kinase activity by stimulating recruitment of MuSK to a preformed oligomer of Lrp4, thereby facilitating MuSK trans-phosphorylation. Lrp4 is essential for synapse formation, and the synaptic defects in *lrp4* mutant mice appear identical to those in *MuSK* mutant mice (Weatherbee, Anderson, & Niswander, 2006).

Interestingly, a pair of proteins with homology to MuSK and Lrp4 are required for clustering AChRs at *C. elegans* neuromuscular synapses: CAM-1, a receptor tyrosine kinase that shares homology with MuSK, and LEV-10, a transmembrane protein that contains an LDLa domain and forms a complex with two secreted proteins, LEV-9 and OIG-4, as well as AChRs (Francis et al., 2005; Gally, Eimer, Richmond, & Bessereau, 2004; Rapti, Richmond, & Bessereau, 2011). However, the kinase activity of CAM-1 is dispensable for synaptic differentiation, pointing to a different mechanism of action than MuSK at vertebrate neuromuscular synapses.

Glutamate, rather than ACh, is the neurotransmitter at *Drosophila* neuromuscular synapses. As such, perhaps it is not surprising that a different class of signals regulate differentiation of these synapses. Although little is known about the initial formation of synapses during embryogenesis in *Drosophila*, two ligands, Wingless and TGF-β, unrelated to Agrin, are involved in the growth of *Drosophila* neuromuscular synapses during larval development. Wingless is secreted by motor neurons and thought to activate Frizzled-2, a Wingless receptor expressed in muscle (Packard et al., 2002), whereas Glass bottom boat, a ligand from the TGF-β family, is synthesized by muscle and thought to activate Wishful thinking, a type II receptor, in motor neurons (Aberle et al., 2002). These signaling mechanisms participate in the regulation of bouton growth and size, which adapt to changes in synaptic activity.

Downstream of MuSK

How does MuSK activation lead to postsynaptic differentiation at vertebrate neuromuscular synapses? Agrin binds Lrp4 and stimulates the rapid phosphorylation of MuSK. The kinase activity of MuSK is essential for Agrin to stimulate clustering and tyrosine phosphorylation of AChRs. Signaling downstream from MuSK depends on phosphorylation of a tyrosine residue (Y553) in the juxtamembrane region of MuSK (Herbst, Avetisova, & Burden, 2002). Phosphorylation of this tyrosine leads to recruitment of Dok-7 (Okada et al., 2006), an adapter protein that further stimulates MuSK and becomes tyrosine phosphorylated (Bergamin, Hallock, Burden, & Hubbard, 2010). Mice lacking Dok-7 have defects identical to *MuSK* mutant mice (Okada et al., 2006), and mutations in human Dok-7 are a major cause of congenital myasthenia (Beeson et al., 2006). Tyrosine phosphorylated Dok-7 recruits Crk and Crk-L, two related adapter proteins that engage additional but unknown downstream signaling pathways essential for synapse formation (Hallock et al., 2010). Rac and Rho, small GTP-binding proteins that regulate actin organization, are activated by Agrin and required for Agrin to stimulate AChR clustering (Weston, Yee, Hod, & Prices, 2000; Weston et al., 2003). Rac acts early and is important for the formation of AChR microclusters, whereas Rho is activated more slowly and is important for the consolidation of microclusters into larger AChR clusters. Moreover, at least one kinase acts downstream from MuSK and is recruited and/or activated by Agrin-activated MuSK, as staurosporine, a protein kinase inhibitor, inhibits Agrin-induced AChR tyrosine phosphorylation and clustering without blocking tyrosine phosphorylation of MuSK. Figure 17.5 illustrates the fundamental signaling components required for forming neuromuscular synapses.

Synapses gradually mature after their initial formation and become fully mature and stable nearly one month after birth in mice. The mechanisms involved in this transition are not well understood, but may involved additional signaling molecules, such as Neuregulin and ErbB receptors, as well as Src-family kinases and cytoskeletal components, such as dystrobrevin (Grady et al., 2000; Smith, Mittaud, Prescott, Fuhrer, & Burden, 2001).

Muscle-Autonomous and Neural-Dependent Patterning of AChR Expression

The preceding sections have described the role of muscle prepatterning and motor neuron-derived signals in regulating the distribution of AChR expression. An initial, spatially restricted pattern of AChR expression is generated in muscle independent of neurally derived Agrin, or indeed of any other neural signal (Fig. 17.2). The mechanisms that are intrinsic to muscle and responsible for establishing

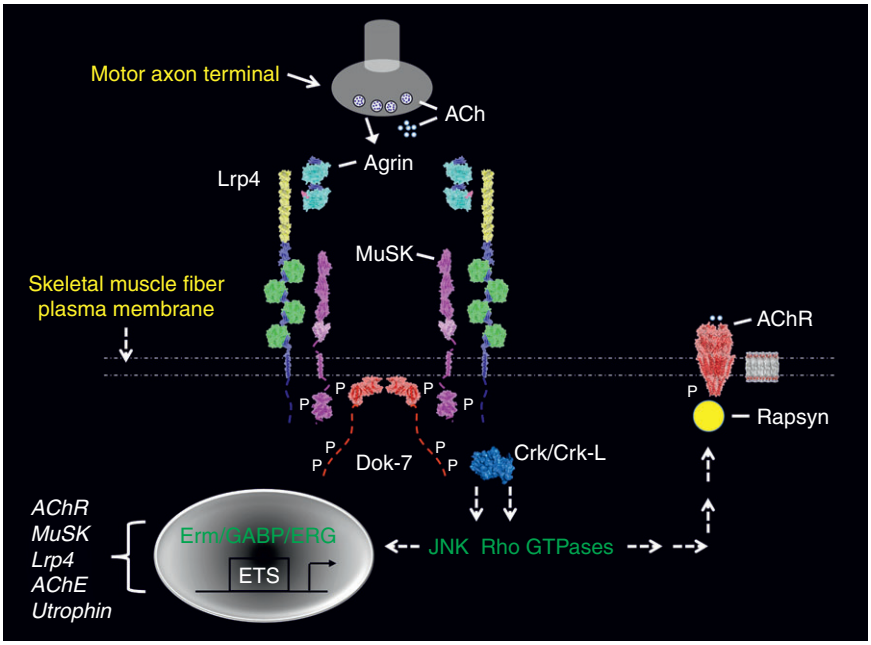


FIGURE 17.5 **Signaling mechanisms for neuromuscular synapse formation.** Agrin, released from motor axon terminals, binds Lrp4, stimulating association between Lrp4 and MuSK and increasing MuSK kinase activity. Phosphorylation of MuSK Y553, in the MuSK juxtamembrane region, causes recruitment and tyrosine phosphorylation of Dok-7, which further stimulates MuSK and recruits Crk and Crk-L adapter proteins to the MuSK/Dok-7 signaling complex. Rac and Rho GTPases function downstream of MuSK/Dok-7/Crk/Crk-L to anchor AChRs at synapses by a poorly understood process that requires Rapsyn, which binds directly to AChRs. The pathway for synapse-specific transcription is less well understood but likely involves JNK-dependent activation of Ets-family transcription factors that stimulate expression of multiple genes encoding synaptic proteins. Synaptic vesicles fuse with the plasma membrane of the nerve terminal, releasing acetylcholine (ACh), which binds to AChRs and causes a conformational change that increases the permeability of the channel to small cations and depolarizes the muscle.

regional differences in muscle in the absence of innervation are not well understood but appear to depend upon the pattern of muscle growth and positive feedback loops that sustain MuSK activation (Kim & Burden, 2008). Because this prepattern is not observed in mice lacking Lrp4, MuSK, or Dok-7, the emergence of this muscle AChR prepattern, nonetheless, coopts some of the same molecules used later during normal synaptogenesis. Additional molecules, however, may be dedicated for prepatterning; for example, zebrafish Wnt11R can function as an alternative direct ligand for MuSK and is important for prepatterning but not synapse formation in zebrafish (Jing, Lefebvre, Gordon, Granato, 2009). Whether Wnts have a role in prepatterning mammalian muscle is not known.

The prepatterned region specifies where motor axons will form synapses. Some synapses form on prepatterned AChR clusters, which become stabilized by Agrin, whereas Agrin also induces new AChR clusters (Flanagan-Steet et al., 2005; Panzer et al., 2006). Thus, the final cohort of synapses on the muscle fiber includes those recognized by motor axons and those induced by motor axons (Arber, Burden, & Harris, 2002; Kummer, Misgeld, & Sanes, 2006) (Fig. 17.2).

Arrival of the nerve, and its attendant signals, converts the AChR prepattern into the more refined pattern of AChR transcription and AChR clustering characteristic of mature synapses. This conversion appears to be dependent on two separable nerve-dependent programs: one program utilizes neurally derived Agrin to maintain AChR expression at nascent synaptic sites, and a second program, which appears to be triggered by acetylcholine, extinguishes AChR clustering at sites that are not stabilized by Agrin/MuSK signaling (Fig. 17.2). These two programs thus ensure the stable expression of AChR clusters selectively at nascent synapses.

Rapsyn is Required for Postsynaptic Differentiation Downstream of Agrin and MuSK

A 43-kDa protein, termed Rapsyn, has an important role in Agrin-mediated signaling. Rapsyn is a myristolated, peripheral membrane protein that is present at 1:1 stoichiometry with AChRs at synaptic sites and interacts directly with AChRs. In addition, Rapsyn links AChRs to components of a subsynaptic, cytoskeletal complex, including Dystroglycan and Dystrophin, Utrophin, Syntrophin, Dystrobrevin, and α-actinin.

Agrin stimulates the clustering of Rapsyn in myotubes grown in cell culture, and clustering of Rapsyn and AChRs occurs coincidentally at developing synapses. Rapsyn is critical for synapse formation, since mice

lacking Rapsyn die within hours after birth and have difficulty moving and breathing (Gautam et al., 1995). Mutations in human Rapsyn are a cause of congenital myasthenia with defects in AChR clustering that lead to muscle weakness (Engel, Ohno, & Sine, 2003). Although the mechanisms by which Rapsyn regulates the clustering of synaptic proteins are poorly understood, the clustering of most synaptic proteins, including AChRs, Utrophin, and Dystroglycan, depends upon Rapsyn. Nonetheless, MuSK is clustered at synapses in *rapsyn* mutant mice, raising the possibility that engagement with Agrin/Lrp4 is sufficient to stabilize MuSK at *rapsyn* mutant synapses. Synapse-specific gene expression (see following) is also normal in *rapsyn* mutant mice, indicating that Rapsyn acts downstream from MuSK to anchor postsynaptic proteins but not in the synapse-specific transcription pathway.

Agrin and MuSK Are Required for Retrograde Signaling and Presynaptic Differentiation

Although pathfinding of motor axons to muscle is normal in mice lacking Agrin or MuSK, *agrin* and *MuSK* mutant mice lack normal nerve terminals. In the mutant mice, branches of the main intramuscular nerve fail to arrest their growth and instead wander aimlessly across the muscle (Fig. 17.4). Because MuSK is expressed in skeletal muscle and not in motor neurons, it seems likely that the aberrant behavior of presynaptic terminals in *MuSK* mutant mice is due to indirect actions of the Agrin/MuSK signaling system. Indeed, restoring MuSK expression selectively in muscle of *MuSK* mutant mice rescues synapse formation and neonatal lethality (Herbst et al., 2002). These results indicate that Agrin, released from nerve terminals, causes the muscle cell, via MuSK activation, to reciprocally provide a recognition signal back to the nerve to indicate that a functional contact has occurred. In response to this putative muscle-derived recognition or adhesion signal, the nerve undergoes presynaptic differentiation and stops growing. Unlike our knowledge of the signals and mechanisms that control postsynaptic differentiation, unraveling the mechanisms that control the differentiation of motor nerve terminals has proved far more challenging and remains one of the glaring gaps in our understanding of neuromuscular synapses. Several muscle-derived signals regulate distinct aspects of presynaptic differentiation: FGFs of the 7/10/22 subfamily, laminin $\beta 2$, collagen IV, and SIRP-α stimulate clustering of synaptic vesicles in cultured motor neurons. Genetic studies support the idea that these signals participate in synaptic differentiation, since neuromuscular synapses do not fully mature in mice that are mutant for FGFR2, laminin $\beta 2$, or collagen IV alpha subunits (Fox et al., 2007). Nonetheless, motor axons terminate and differentiate to a considerable extent in the absence of these signaling components, indicating that additional, yet to be identified retrograde organizers regulate earlier steps in presynaptic differentiation.

Certain Genes Are Expressed Selectively in Synaptic Nuclei of Myofibers

Like AChR protein, mRNAs encoding the different AChR subunits (α, β, γ or ε and δ) are concentrated at synaptic sites. Studies with transgenic mice that harbor gene fusions between regulatory regions of *AChR* subunit genes and reporter genes have shown that *AChR* genes are transcribed selectively in myofiber nuclei near the synaptic site (Fig. 17.6). Thus, localized transcription of *AChR* genes in synaptic nuclei is responsible, at least in part, for the accumulation of *AChR* mRNA at synaptic sites. This pathway is important for ensuring that AChRs are expressed at the required density in the postsynaptic membrane, since defects in synapse-specific gene expression of the *AChR* ε subunit gene are a cause for congenital myasthenia (see later).

Like *AChR* subunit genes, the *utrophin* gene is transcribed selectively in synaptic nuclei, resulting in an accumulation of *utrophin* mRNA and protein at synaptic sites. mRNAs encoding Rapsyn, NCAM, MuSK, sodium channels, the catalytic subunit of AChE, LL5beta, and CD24 are also concentrated in the synaptic region of skeletal myofibers, raising the possibility that these genes are likewise transcribed preferentially in synaptic nuclei (Fig. 17.5).

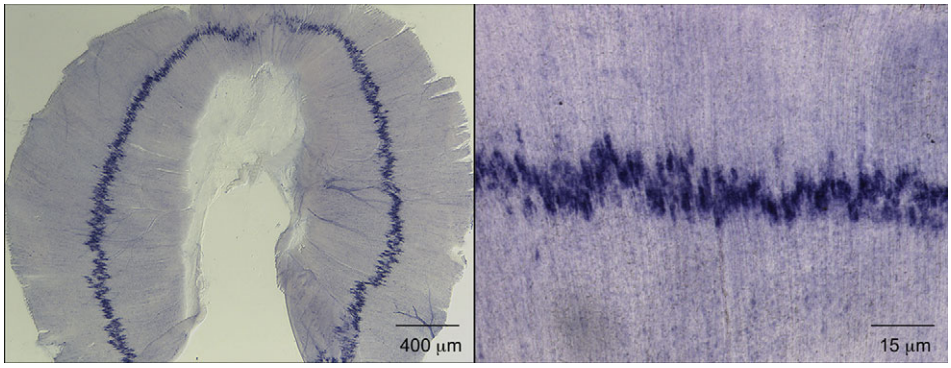


FIGURE 17.6 *Acetylcholine receptor* (*AChR*) genes are expressed selectively by myofiber synaptic nuclei, leading to enrichment of *AChR* mRNA in the central, synaptic region of the muscle. A whole mount of a mouse diaphragm muscle was processed for *in situ* hybridization to reveal the pattern of *AChR* δ mRNA expression. *AChR* δ mRNA expression is concentrated in a narrow band in the central, synaptic region of the muscle.

Thus, synapse-specific transcription may be a common and important mechanism for expressing and localizing a variety of gene products at high levels at the neuromuscular synapse.

The neural signals that regulate synapse-specific transcription remain elusive. Neuregulin-1, which stimulates tyrosine phosphorylation of ErbB2, a member of the EGF receptor family, stimulates *AChR* gene expression in cultured muscle cells, but neuromuscular synapses form normally during development in mice lacking ErbBs in skeletal muscle, as well as in mice lacking Neuregulin-1 in motor neurons and skeletal muscle (Escher et al., 2005; Jaworski & Burden, 2006). Synapse-specific transcription *in vivo*, however, requires Agrin, Lrp4, and MuSK, indicating that this signaling pathway has a critical role in regulating synaptic gene expression.

Ets Domain Transcription Factors Regulate Synapse-Specific Transcription

A binding site for Ets domain proteins in the *AChR* delta subunit gene is critical for synapse-specific gene expression in mice. Importantly, mutation of an Ets-binding site in the human *AChR* ε subunit gene leads to congenital myasthenia, due to decreased AChR expression. This Ets site can bind GABP, a complex containing GABPα, an Ets protein, and GABPβ, a protein that lacks an Ets domain but dimerizes with GABPα, suggesting that GABP may be a transcriptional regulator that stimulates transcription of *AChR* genes in synaptic nuclei (Fig. 17.5). In addition, the gene encoding another Ets domain protein, Erm, is expressed selectively by myofiber synaptic nuclei (Hippenmeyer et al., 2002), suggesting that Erm may also have a role in synapse-specific expression. Genetic analysis of *GABP* and *erm* mutant mice indicates that Erm has a critical role, whereas GABP has a more subsidiary role in synapse-specific transcription.

Electrical Activity Regulates Gene Expression

Changes in the pattern of muscle electrical activity have an important role in regulating the electrophysiological and structural properties of muscle, as well as the ability of motor axons to innervate muscle. The expression of several genes, including *AChR* genes, is repressed by electrical activity, and this repression, together with focal activation in synaptic nuclei, as described earlier, contributes to the disparate levels of *AChR* expression in synaptic and nonsynaptic regions of the muscle. A binding site for myogenic bHLH transcription factors, or E box, in the proximal promoter of *AChR* subunit genes is essential for electrical activity-dependent transcription, since transgenes containing a mutation in this E box, unlike wild-type transgenes, are not induced following denervation. These results suggest that electrical activity decreases the level and/or activity of E box binding proteins, notably Myogenin, leading to decreased *AChR* expression. The transcriptional mechanisms that control expression of *myogenin* in an activity-dependent manner are just beginning to be explored. Electrical activity induces expression of two transcriptional repressors, HDAC-9 (aka MITR) and Dach2, which act to repress *myogenin* expression (Mejat et al., 2005; Tang and Goldman 2006). A reduction in myofiber electrical activity leads to a reduction in *HDAC-9* and *Dach2* expression, causing an elevation in *myogenin* expression. Thus, the mechanisms that couple changes in electrical activity to changes in *HDAC-9* and *Dach2* expression are critical for regulating genes, such as *AChR* genes, which are dependent upon Myogenin. The mechanisms that regulate *HDAC-9* and *Dach2* expression are not understood, but the involvement of multiple protein kinases, including PKA, CAM kinase II, and PKC, has been suggested.

Summary

Signals exchanged at developing synapses ensure that differentiated presynaptic terminals are aligned precisely with a highly specialized postsynaptic membrane. Analysis of mice lacking innervation demonstrates that a coarse pattern of postsynaptic differentiation is present prior to innervation and that neuronal signals, including Agrin, both maintain and induce postsynaptic differentiation at sites of nerve–muscle contact (Fig. 17.2). We are gaining a better understanding of the pathways that act downstream from MuSK to control postsynaptic differentiation (Fig. 17.5), but we remain poorly informed about the signals and pathways that control presynaptic differentiation.

SYNAPSE FORMATION IN THE CENTRAL NERVOUS SYSTEM

Compared to synapse formation between neurons and muscle cells (at the neuromuscular synapse), we currently know relatively little about how neuron-to-neuron synapses in the central nervous system are put together. The reason for this relative lack of knowledge is the complexity of the problem. CNS synapses are much smaller (about one-hundredth the size of a neuromuscular synapse), a single neuron typically receives hundreds to thousands of synapses, and the synaptic inputs are derived from multiple different synaptic partners. Finally, these synapses employ not just one type but a variety of neurotransmitters. Some of these neurotransmitters are excitatory and depolarize the postsynaptic partner, and some are inhibitory and lead to hyperpolarization. The placement of excitatory and inhibitory synapses on the

postsynaptic cell is a critical parameter for integration of these counteracting inputs. Synaptic connectivity in the CNS is not static but undergoes extensive plastic changes. In response to experience, synaptic connections and the function of neuronal circuits are modified. All of these CNS synapse properties pose intriguing questions regarding the synthesis of synaptic components, their targeting and assembly at functionally distinct synaptic sites, and the recognition of the appropriate synaptic partners amongst an array of candidate partners.

Despite these differences between peripheral and CNS synapses, there are several common principles that apply to both systems: (1) Like neuromuscular synapses, CNS synapses are tri-partite structures. Highly specialized presynaptic release sites are closely apposed to postsynaptic specializations that concentrate neurotransmitter receptors (Fig. 17.7). This pair of synaptic partners is lined by astroglial processes that ensheath the synaptic cleft. (2) Cells contain most synaptic proteins before synapse formation and bidirectional signaling organizes these components at emerging synaptic contacts. (3) Synapses undergo extensive functional maturation. (4) During development, an exuberant number of synaptic contacts are formed, some of which later undergo elimination.

The following section focuses on the formation of CNS synapses in vertebrates. Much research has been on glutamatergic synapses—excitatory synapses using the neurotransmitter glutamate that are the most abundant in the CNS. We will also highlight important insights gained in the analysis of synapses using other neurotransmitters. We will start by discussing the time-course of *morphological rearrangements* of synapse formation and then examine emerging evidence on the mechanisms controlling *recruitment and assembly* of pre- and postsynaptic components. Finally, we will describe some candidate signals that may contribute to the *selectivity* of synaptic wiring during development.

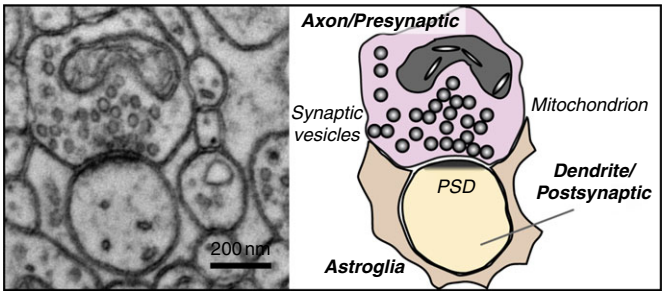


FIGURE 17.7 **Synapses are tripartite junctions.** The electron micrograph shows a parallel fiber synapse in the mouse cerebellum, an example of a glutamatergic synapse. The presynaptic terminal contains a mitochondrion and clusters of synaptic vesicles throughout the terminal. A few vesicles are morphologically docked at the active zone opposite the postsynaptic density. The postsynaptic side is a dendritic spine that in the cross section appears like a circular structure. The electron-dense material opposite the vesicle release sites is the postsynaptic density (PSD). Synaptic cleft and dendritic spine are closely ensheathed by an astroglia process (image courtesy of Stéphane Baudouin).

Initiation of Axon-Target Interaction

Prior to synapse formation, neurons already express both pre- and postsynaptic proteins. For example, growing axons and growth cones contain synaptic vesicles and release neurotransmitter. Similarly, target cells (that will later become the postsynaptic side) express functional neurotransmitter receptors on their surface. Cell–cell contact then is essential for the concentration and juxtaposition of pre- and postsynaptic components in opposing membranes and for achieving the mature complement of both pre- and postsynaptic proteins.

Many initial contacts between synaptic partners are established through filopodia, tiny actin-based extensions that rapidly protrude and retract from a cell. Importantly, filopodia extend from both axons and dendrites and increase the extracellular space sampled by a cell during development. Imaging studies, both in culture and *in vivo*, indicate that when filopodia contact potential synaptic partners, they can lose their motility and become stabilized. These contacts are then thought to transform into synaptic structures. Indeed, synaptic specializations have been seen on filopodia themselves, and neurotransmitter release from axons is likely to play an important role for the regulation of filopodial dynamics.

A further important realization from imaging studies has been that the assembly of synaptic structures occurs rapidly. The time required from initial contact to establishment of a functional synapse is only in the range of 1–2 hours, with the formation of a presynaptic terminal occurring in only 10–20 minutes and the recruitment of postsynaptic neurotransmitter receptors taking place in approximately 90 minutes after cell–cell contact (Garner, Zhai, Gundelfinger, & Ziv, 2002). How is the entire machinery for regulated vesicular release recruited within such a short time frame? This timecourse is plausible if one considers that most synaptic components are "ready-to-go" before contact and primarily need to be redistributed (rather than newly synthesized). Still, the speed of synaptic assembly is remarkable. One intriguing model poses that presynaptic components are recruited in preassembled units (Fig. 17.8). Synaptic vesicles move in axons before synapse formation in grape-like clusters. Similarly, multiple active zone components such as voltage-gated calcium channels and syntaxin can be found prepackaged in so-called "piccolo-bassoon transport vesicles" (PTVs) (Garner et al., 2002). PTVs are vesicles named after the presynaptic active zone proteins piccolo and bassoon that are closely associated with these carriers that appear as 80-nm-diameter dense core vesicles in electron microscopy. PTVs could be rapidly

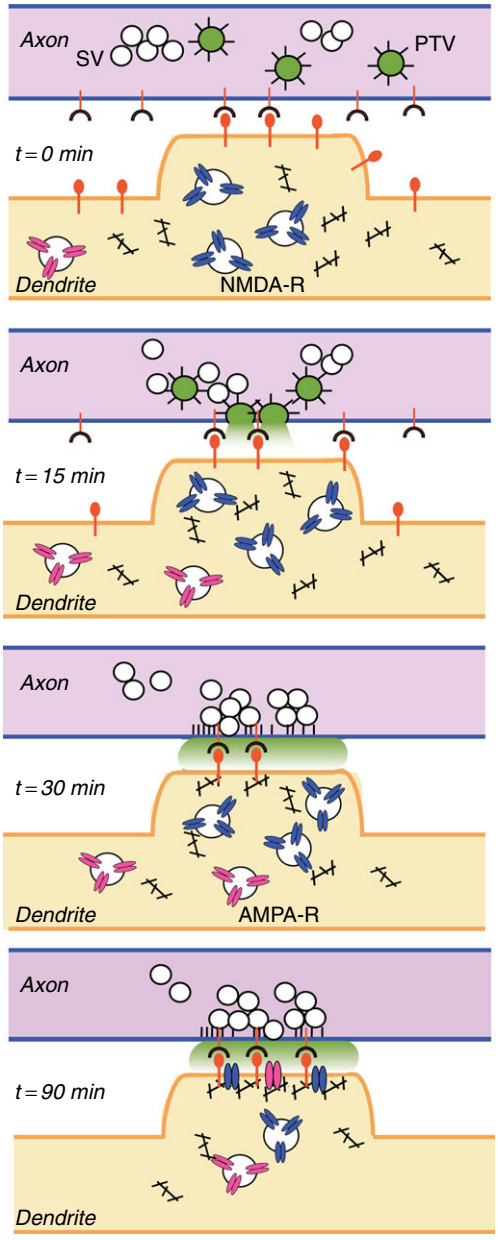


FIGURE 17.8 **Hypothetical model for the assembly of central synapses.** Before synapses formation, axons and dendrites contain synaptic vesicles (SV), active zone components, and neurotransmitter receptors, respectively. Cell–cell contact is initiated via cell surface adhesion and/or signaling molecules (red, $t = 0$ min). Contact triggers fusion of piccolo-bassoon transport vesicles (PTV, green) with the plasma membrane ($t = 15$ min), which leads to the deposition of active zone material (black spikes). Subsequently, synaptic vesicles (clear circles) are recruited to the cell–cell contact sites ($t = 30$ min), and postsynaptic scaffolding molecules (black), NMDA-type (blue), and AMPA-type neurotransmitter receptors (red) are delivered sequentially to the maturing postsynaptic membrane. *Adapted from Garner et al. (2002).*

captured and inserted in response to a target-derived signal, thereby greatly simplifying the process of presynaptic assembly (Fig. 17.8). Subsequently, neurotransmitter receptors are recruited to the postsynaptic domain. Excitatory synapses using the transmitter glutamate contain the ionotropic NMDA- and AMPA-type receptors (see Chapter 9). After cell–cell contact, NMDA- and AMPA-receptors are recruited to synapses in a stepwise manner, with an initial accumulation of NMDA receptors at postsynaptic sites and the subsequent addition of AMPA receptors (Fig. 17.8). Candidate signals for the neurotransmitter receptor accumulation at synapses and for the stabilization of filopodial contacts between pre- and postsynaptic partners are discussed next.

Molecular Signals for Synaptic Development in the CNS

What are the signals that mediate the morphological and functional differentiation of pre- and postsynaptic structures? It turns out that several transsynaptic signals "singlehandedly" induce a substantial degree of synaptic organization. Based on this ability to drive synapse assembly, these signals have been termed "synapse-organizing" or "synaptogenic" factors. Synaptogenic proteins are employed in multiple ways to drive synapse organization: First, target cells (the future postsynaptic partner) secrete growth factors that instruct presynaptic differentiation in axons. We refer to these signals as *retrograde* organizers of synapses. Second, axons (the future presynaptic partner) release signals that drive the postsynaptic accumulation of neurotransmitter receptors, much like Agrin at the neuromuscular junction. These signals are referred to as *anterograde* organizers. Third, several adhesion and signaling complexes bridge the synaptic cleft and simultaneously recruit pre- and postsynaptic components; consequently, these are considered *bidirectional* organizers of synapses. Finally, synaptic astroglia also contribute signals that promote formation and maturation of central synapses. Examples for each of these four different classes of synapse-organizing mechanisms are discussed following.

Morphogens as Retrograde Signals for Synaptic Organization

Morphogens are secreted growth factors that direct cell fate decisions during embryonic development. Components of many morphogen signaling cascades continue to be expressed in the postnatal brain, indicating that they may govern additional developmental processes after the completion of cell fate specification. For example, Wnts and fibroblast growth factors (FGFs) exhibit important, previously unanticipated signaling functions in axon guidance and synapse formation—essentially being "recycled" for morphogenetic programs in a different developmental context during postnatal development.

At CNS synapses, Wnts and FGFs both act as retrograde signals. This means they are released from synaptic target cells and promote presynaptic differentiation

in afferents. Synaptic signaling functions were first discovered in the mouse cerebellum for Wnt7a, a vertebrate homologue of the *Drosophila melanogaster* wingless gene (Fig. 17.9). Wnt7a is released from cerebellar granule cells and promotes synaptic vesicle accumulation in mossy fiber axons, a class of afferent projections derived from the brainstem and the spinal cord (Hall, Lucas, & Salinas, 2000). Presynaptic vesicle recruitment in response to Wnt7a is most likely regulated through local modification of microtubule branching and remodeling. Wnt7a signaling is mediated through the axonal receptor frizzled, a receptor that also plays a central role in Wnt signaling during early development. Frizzled acts through the signaling adapter disheveled-1 (Dvl-1) and the cytoplasmic serine/threonine kinase GSK3beta. In turn, GSK3beta regulates microtubule-associated proteins (MAP1B and adenoma polyposis complex, APC) by phosphorylation to shape a specialized microtubule network in growth cones and axonal branches.

Fibroblast growth factors comprise a large family of secreted growth factors expressed in neuronal and nonneuronal cells. The closely related isoforms FGF22, FGF7, and FGF10 are target-derived signals that contribute to presynaptic development (Umemori, Linhoff, Ornitz, & Sanes, 2004). In the absence of FGF-signaling presynaptic vesicle accumulation in axons is reduced. Notably, neither Wnts nor FGFs organize the assembly of active zones, indicating that they preferentially promote the expansion of the presynaptic vesicle pools. FGF signaling is mediated through an axonal FGF receptor complex, a receptor tyrosine kinase that couples to the PI3-kinase and Ras-MAP-kinase pathway. Similar to Wnt7a, FGFs alter cytoskeletal properties and promote axonal branching. These morphological rearrangements shape the presynaptic partner and facilitate the recruitment of appropriate organelles in response to additional transsynaptic synaptogenic signals.

Transsynaptic Adhesion Complexes as Bidirectional Synapse Organizers

Adhesion molecules are central organizers of specialized cellular junctions in all tissues. Importantly, several proteins with adhesive properties have emerged as regulators with synaptogenic activity (Fig. 17.9). One class of these factors are the neuroligins, postsynaptic adhesion molecules that can bind to neurexins, a family of receptors present on axons (Sudhof, 2008). Neuroligin-neurexin interactions mediate heterophilic adhesion across the synaptic cleft, thereby directly connecting the pre- and postsynaptic compartments. Heterologous expression of neuroligin in nonneuronal cells induces focal accumulations of neurexins and synaptic vesicle proteins in contacting axons (Scheiffele, Fan, Choih, Fetter, & Serafini, 2000). The neuroligin-induced presynaptic structures are capable of vesicle exo- and endocytosis upon depolarization, indicating that not only synaptic vesicles but the entire machinery for regulated exocytosis is recruited to these nascent presynaptic terminals. Interestingly, neurexins also serve as a synaptic receptor for another postsynaptic adhesion molecule called LRRTM2 (leucine-rich repeat protein 2). LRRTM2 drives presynaptic differentiation through direct interaction with axonal neurexins, therefore, providing an alternate neurexin ligand to neuroligins (de Wit et al., 2009). The neuroligin-neurexin and LRRTM2-neurexin transsynaptic complexes are part of a growing list of proteins with synaptogenic functions. One additional notable example is the postsynaptic receptor TrkC (Takahashi et al., 2011). TrkC is a well-known receptor for neurotrophins that regulates neuronal survival (see Chapter 19). Surprisingly, at synapses TrkC engages in adhesive interactions with presynaptic receptor protein tyrosine phosphatases (RPTPs). These interactions promote recruitment of synaptic vesicles to presynaptic sites and the maturation of the postsynaptic structures. Notably, the synaptogenic function of TrkC

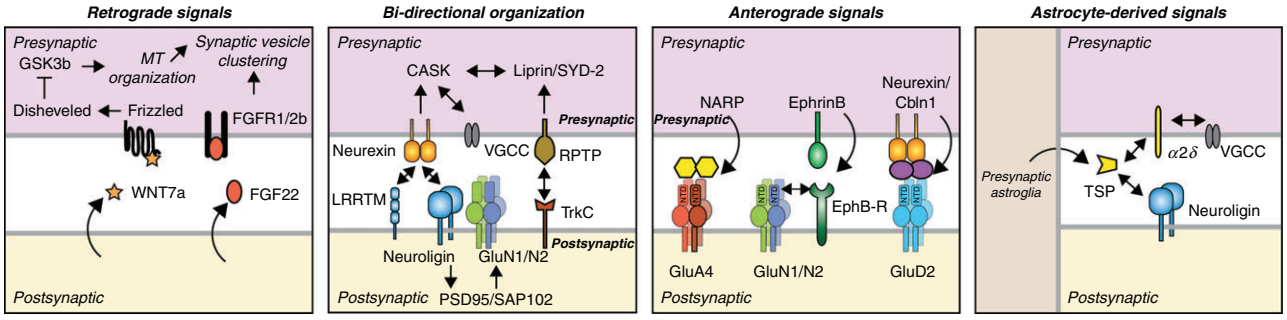


FIGURE 17.9 **Synapse-organizing signals at CNS synapses.** Different modes of synaptic organizing signals exchanged between pre- and postsynaptic partners and glial cells are illustrated. Wnt7a and FGF22 are retrograde signals derived from the postsynaptic target cell. Transsynaptic adhesion complexes link to scaffolding molecules which interact with presynaptic voltage-gated calcium channels (VGCC) or postsynaptic NMDA-receptors (GluN1/N2), respectively. Several anterograde signals from the presynaptic cell exert their synapse-organizing functions via direct interaction with the amino-terminal-domain (NTD) of glutamate receptors. Perisynaptic astroglia release synapse-organizing molecules such as thrombospondin (TSP). See text for further discussion.

does not require the presence of neurotrophins or the TrkC kinase activity. This suggests that in addition to its role in cell survival, TrkC has evolved functions that resemble an adhesion molecule.

How do transsynaptic adhesion complexes promote the assembly of synaptic junctions? The synaptogenic functions of multiple transsynaptic adhesion complexes appear to converge onto a common platform of presynaptic signal mediators. Neurexins couple via PDZ-domain interactions to a group of multidomain scaffolding molecules including Lin-2/CASK and Lin-10/Mint. RPTPs (the presynaptic receptors for TrkC) couple to liprins, which in turn interact with Lin-2/CASK. This presynaptic scaffolding complex might link synaptic adhesion molecules to voltage-gated calcium channels—the key calcium source for regulated secretion at the active zone (Missler et al., 2003). Currently, evidence for essential roles for any of these scaffolding molecules in presynaptic assembly in vertebrates is lacking. However, analysis of invertebrate model systems highlighted that these proteins are indeed key regulators of presynaptic assembly. *C. elegans* liprin has been recovered as a mutation called *synapse-defective-2* in a genetic screen for mutant nematodes with altered presynaptic differentiation (Jin, 2002). In *synapse-defective-2* mutant worms active zones are unusually lengthened and synaptic vesicles are diffusely localized along presynaptic sites. By analogy with the RPTP-liprin interactions observed in vertebrates, SYD-2 acts downstream of the receptor tyrosine phosphatase called PTP-3A. Exploring the function of liprins/Syd-2 in vertebrates is complicated by the fact that multiple independent genes encode liprin isoforms. However, expression, localization and biochemical interactions of liprins are consistent with a role in presynaptic assembly.

The concept of submembrane scaffolds organized by synaptic adhesion complexes not only applies to presynaptic organization but such scaffolds may also underlie the concentration of postsynaptic neurotransmitter receptors. In analogy to AChR clustering at the neuromuscular synapse, transsynaptic signals recruit Rapsyn-like cytoplasmic proteins that provide a scaffold lining the postsynaptic membrane. Different forms of such scaffolds selectively retain postsynaptic neurotransmitter receptors at the appropriate sites. Biochemical studies have identified a vast array of such scaffolding proteins. These proteins are characterized by combinations of protein–protein interaction domains, making them well suited to cross-link postsynaptic receptors into larger complexes (Kim & Sheng, 2004). Neuroligins bind, via PDZ-domain interactions, to SAP102 and PSD-95, major components of the glutamatergic postsynaptic scaffold. Aggregation of neuroligins results in the recruitment of the scaffolding proteins and NMDA-type glutamate receptors (Chih, Engleman, & Scheiffele 2005; Graf, Zhang, Jin, Linhoff, & Craig, 2004). Mice lacking neuroligin proteins show a subtle reduction in synapse density and a significant loss of postsynaptic neurotransmitter receptors from synapses (Varoqueaux et al., 2006). These studies highlight the coupling of neurotransmitter receptors to adhesion molecules and a potential role for scaffolding molecules in this process. The importance of these interactions is further emphasized by the association of mutations in scaffolding and adhesion molecules with autism-spectrum disorders in human patients (Box 17.2).

The concept that scaffolding molecules contribute to the clustering of postsynaptic neurotransmitter receptors applies to both excitatory and inhibitory synapses. Importantly, these synapses differ not only in their presynaptic neurotransmitters and postsynaptic neurotransmitter receptors but also in their scaffolding proteins. In fact, scaffolding molecules may ensure that the appropriate neurotransmitter receptors are selectively recruited to excitatory and inhibitory synapses, respectively. The scaffolding protein gephyrin represents one of the best-understood examples for the scaffold-mediated postsynaptic receptor clustering at inhibitory synapses. Gephyrin

BOX 17.2

SYNAPTIC TARGETS IN NEURODEVELOPMENTAL DISORDERS

Human genetic studies have uncovered genetic risk factors for autism-spectrum disorders. Autism is characterized by impairments in social interactions and communication. Patients exhibit a spectrum of restricted interests, ritualized behaviors, and stereotyped, repetitive movements. Autism is considered neurodevelopmental in origin as the defects emerge in the first years of life during the developmental time period of synapse formation and remodeling. Multiple genetic alterations (mutations or copy number variations) identified in patients encode proteins with synaptic function. For example, genes encoding the neurexin-neuroligin adhesion complex have emerged as autism candidate genes, including Nrxn1, Nlgn1, Nlgn3, Nlgn4, and the associated scaffolding molecules Shank2 and Shank3. Mutant mice recapitulating these mutations exhibit behavioral alterations that mimic core defects of autism and serve as models for developing therapeutic strategies.

Steven J. Burden and Peter Scheiffele

is a cytoplasmic protein that associates with the glycine receptor and is essential for receptor clustering at synaptic sites (Kneussel & Betz, 2000). Gephyrin is well equipped to serve as a scaffolding protein that links receptors to the cytoskeleton; it binds with high affinity to polymerized tubulin, the actin-binding protein profilin, and the lipid-binding protein collybistin. Thereby, gephyrin may link membrane lipids, neurotransmitter receptors, and cytoskeletal elements to stabilize postsynaptic specializations. Moreover, the gephyrin-collybistin complex is well positioned to recruit GABA-A-receptors in response to transsynaptic adhesion through a mechanism analogous to the scaffolding protein interactions at glutamatergic synapses. Neuroligin-2, a member of the neuroligin protein family that is concentrated at GABAergic synapses, activates collybistin, which in turn tethers gephyrin at the postsynaptic membrane. These gephyrin clusters then retain glycine and GABA-A-receptors at postsynaptic sites (Poulopoulos et al., 2009). In the hippocampus of neuroligin-2 knockout mice, the perisomatic accumulation of gephyrin and GABA-A receptors is reduced, which highlights an essential role for neuroligin-2 in nucleating a cytoplasmic protein platform of GABAergic synapses. In summary, multiple transsynaptic adhesion systems provide bidirectional signals that cooperate in the organization of pre- and postsynaptic structures.

Postsynaptic Neurotransmitter Receptors as Effectors of Anterograde Signaling

While cytoplasmic scaffolding protein interactions provide one important pathway for the clustering of postsynaptic neurotransmitter receptors, direct extracellular (scaffold-independent) interactions with neurotransmitter receptors themselves have emerged as a second mechanism (Fig. 17.9). Ionotropic glutamate receptors are tetramers, generally composed of two molecules each of two different subunits. In addition to ligand-binding and pore-forming domains, glutamate receptor subunits contain an extracellular amino-terminal domain (NTD). Proteins that directly bind to these NTDs serve as anterograde signals for the synaptic recruitment of glutamate receptors. The neuronal pentraxins were the first such anterograde signals to be identified. Pentraxins (NARP and NP1) are released from axons and interact directly with the NTD of AMPA-receptor (GluA) subunits. This interaction is sufficient to drive the synaptic aggregation of receptors at glutamatergic postsynaptic sites (O'Brien et al., 1999). Importantly, NTD-mediated interactions for postsynaptic differentiation are not unique to AMPA-type glutamate receptors but are emerging as a general principle. The NTD of the NMDA-receptor subunit GluN1 mediates direct interactions with EphB2, a member of the Eph-receptor family of receptor tyrosine kinases (Dalva et al., 2000). EphB-receptors form transsynaptic complexes with their ligands, called ephrinBs. These interactions nucleate a tripartite complex consisting of ephrinB, EphB-receptor, and NMDA-receptors, providing an anterograde signal for NMDA-receptor recruitment to synapses. Notably, EphB-receptors had been initially recognized for their critical roles in axon guidance and topographic mapping. The function of the EphrinB-EphB-receptor system at synapses highlights additional functions. In fact, other axon guidance molecules were recently identified as regulators of synapse formation. In many cases the signaling read-out downstream of these receptors varies between the different steps of neural development: some receptors that mediate repulsive signaling during axon guidance (such as EphB) have adhesive and stabilizing functions at synapses. How this functional switch is achieved is currently unknown.

Arguably, the most remarkable example for direct neurotransmitter receptor interactions in synaptic differentiation is provided by the GluD2 protein. GluD2 forms homo-tetramers and shares the same domain organization with AMPA- and NMDA-type receptors. However, owing to amino acid alterations in the receptor pore, GluD2 does not appear to mediate any currents. GluD2 is selectively expressed in cerebellar Purkinje cells where it is critical for synaptic connectivity and function. Mice lacking GluD2 suffer from cerebellar ataxia—that is, defects in motor coordination and balance. Parallel fiber synapses that form between cerebellar granule cells and Purkinje cells detach, indicating a severe weakening of transsynaptic adhesion. The GluD2 NTD stabilizes synapses through the presynaptic ligand cerebellin-1 (Cbln1), a small, secreted protein (Uemura et al., 2010). Whereas the GluD2 receptor is accumulated in Purkinje cell dendritic spines, Cbln1 is secreted from cerebellar granule cells and concentrates at parallel fiber synapses with Purkinje cells. Interestingly, Cbln1 binds to presynaptic neurexins (Fig. 17.9)—that is, the same presynaptic receptor that also transduces synaptogenic activities of postsynaptic neuroligins and LRRTM proteins (see above). In summary, Neurexin, Cbln1, and GluD2 form a tripartite complex that promotes synapse formation via the amino-terminal domain of GluD2. GluD2 is expressed exclusively in cerebellar Purkinje cells but neurexins and Cbln molecules are found more broadly in the nervous system and are likely to contribute to differentiation of other synapses as well. In fact, the cell type–specific function of the three anterograde synapse-organizing systems described in this paragraph highlight an important principle for CNS synapse formation: synapses formed between different neuronal partners have evolved a diverse range of signaling modules that control their assembly. The Agrin-Lrp4/MuSK complex is a universal signal for stabilization of all neuromuscular synapses. By contrast, signaling systems in the CNS are much more diverse and different synapses assemble by different mechanistic rules.

Glial-Derived Signals for Synapse Development

All transsynaptic signals we have discussed so far are exchanged directly between the pre- and postsynaptic partners. However, synapses are tripartite structures in which glial cells assume highly specialized morphologies and functions (Fig. 17.7). Importantly, some synaptic differentiation factors are derived from glial cells and coordinate the assembly of pre- and postsynaptic structures. Cultured astrocytes greatly enhance the number of synapses formed by cultured purified retinal ganglion cells (Eroglu & Barres, 2010). Astrocytes secrete multiple factors to control various aspects of synapse formation. One of these factors is thrombospondin (TSP), a large secreted proteoglycan molecule (Fig. 17.9). TSP induces morphologically defined synapses in cultured neurons, and mutant mice with perturbed TSP expression have fewer synapses. This suggests that this astrocyte-derived protein is required for synapse formation in the intact nervous system. The activity of TSP is mediated by interactions with both pre- as well as postsynaptic receptors. A presynaptic TSP receptor is the protein alpha-2-delta, a transmembrane protein associated with voltage-dependent calcium channels. On the postsynaptic side, TSP interacts with neuroligin-1, which in turn couples to NMDA-type glutamate receptors. This example highlights that glial cells play important roles in determining the formation and function of synapses during development. Additional glial-derived signals for synapse formation are likely to exist and remain to be discovered.

Surface Recognition Molecules in Synaptic Specificity

Understanding the molecular mechanisms of synaptic specificity is arguably the most exciting unresolved question in the field of CNS synaptogenesis. A leading model is that selective transsynaptic interactions are achieved by an adhesive code. Such a code could be generated by corresponding attractive cues in pre- and postsynaptic cells, matching appropriate partners through adhesive interactions. Such attractive cues for selective wiring are complemented by repulsive signals that limit the interaction between inappropriate partners. In the following, we will discuss examples for both mechanisms.

Cadherins are the first proteins proposed to function as adhesive synaptic specificity factors. Cadherins are a family of homophilic adhesion molecules that are major regulators of cell morphology and function in all tissues. In the CNS, cadherins are concentrated at synapses. Since there are about 20 different subtypes of classical cadherins these proteins could encode selective interactions between synaptic partners. In many brain areas specific cadherin isoforms are selectively expressed in corresponding pre- and postsynaptic partners, making them excellent candidates to impose specificity on the synapse formation process. This hypothesis has been tested in the hippocampal mossy fiber projection in mice (Fig. 17.10). In the hippocampal circuit, mossy fiber axons derived from the dentate granule cells (DG) form large glomerular synapses selectively on CA3 pyramidal cells. Interestingly, preferential synapse formation between DG and CA3 cells is maintained, even in dissociated cultures of hippocampal neurons where spatial constraints imposed by the hippocampal architecture are absent. Therefore, it is indeed cellular properties and not cell location in the neuropil that are important for selective synaptic interactions. DG and CA3 cells both express the same cadherin isoform (cadherin-9). In the absence of cadherin-9, DG-CA3 synapses are under-developed and form in lower numbers (Williams, et al., 2011). These experiments provide a first glimpse of critical roles for cadherin-mediated adhesion in selective synapse formation. Given that multiple different cadherin isoforms are found throughout different brain regions and neuronal circuits, analogous cadherin-mediated interactions may contribute to further cases of synaptic specificity.

Beside cadherins, additional homophilic adhesion molecules have been implicated in synaptic specificity. The vertebrate retina is a particularly useful system for the analysis of selective wiring due to its stereotyped organization (Fig. 17.10). Amacrine and bipolar cells elaborate neurite arborizations in the inner plexiform layer, a laminar region where synapses on retinal ganglion cell dendrites are established. The inner plexiform layer is divided into so-called sublaminae defined by the restriction of neurites from distinct amacrine and bipolar cells and their retinal ganglion cell targets. Cells elaborating neurites and synapses in specific sublaminae differ in their activation upon visual stimuli (ON and OFF cells; see Chapter 9 for details). A family of immunoglobulin domain proteins, called Sidekicks and the closely related DSCAMs, are expressed in synaptic partners (Yamagata & Sanes, 2008). Sidekicks and DSCAMs mediate homophilic adhesion and direct axo-dendritic interactions of specific pairs of amacrine and retinal ganglion cells to specific sublaminae (Fig. 17.10). Therefore, similar to the hippocampus, a substantial degree of laminar specificity can be achieved through adhesive interactions between pre- and postsynaptic partners. However, not all aspects of laminar specificity rely on attractive interactions. Repulsion from inappropriate targets has emerged as a second fundamental mechanism for achieving synaptic specificity. A compelling example for this mechanism is observed for the laminar stratification of amacrine cell neurites (Matsuoka et al., 2011). A subclass of amacrine cells (containing the neuromodulator dopamine or calbindin) elaborates neurites and synapses predominantly in the S1 and S2 sublaminae of the IPL. This restriction is achieved by

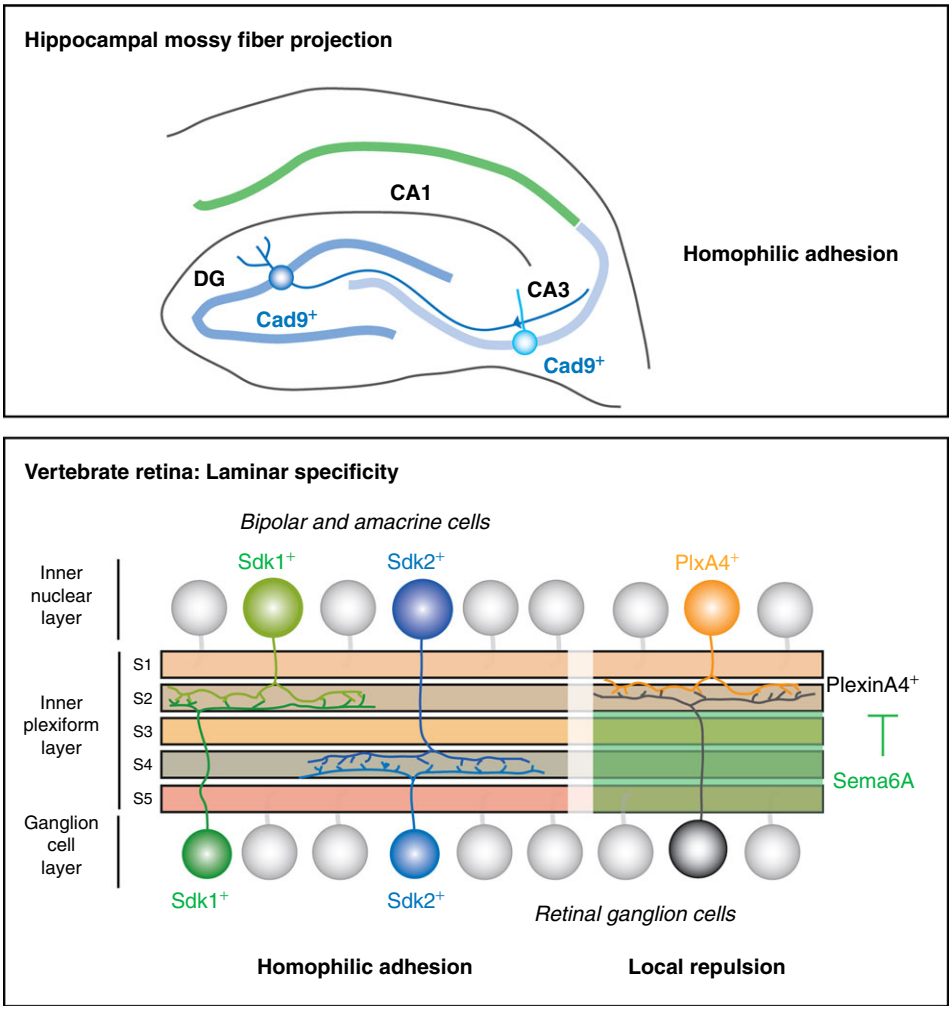


FIGURE 17.10 **Model mechanisms for synaptic specificity.** (A) In the hippocampus, dentate granule cells (DG) form synapses selectively on pyramidal cells in the CA3 region. Pre- and postsynaptic partners express the same homophilic adhesion molecule (cadherin-9) and thereby preferentially form synapses with each other. (B) In the vertebrate retina, the analysis of laminar specificity of neurite elaboration is facilitated by stereotyped layered organization of the inner plexiform layer (sub-laminae S1-S5). In this system, attractive and repulsive signals have been identified which guide synaptic specificity. Subsets of amacrine and bipolar cells in the inner nuclear layer contain different isoforms of the protein Sidekick (Sdk1 or Sdk2). Through homophilic adhesion, these cells form lamina-specific synaptic connections on the dendrites of Sdk1 or Sdk2-positive ganglion cells, respectively. In the same system, local repulsion is used as an additional specificity mechanism. The neurites elaborated by PlexinA4-positive amacrine cell ($PlxA4^+$) are repelled by semaphorin6A (Sema6A) expressed in sub-laminae S3-S5. Thereby, $PlxA4^+$ neurites are restricted to sub-laminae S1 and S2. *Figures adapted from Williams, de Wit, and Ghosh (2010) and Yamagata and Sanes (2008).*

repulsive signals transduced through the Plexin-Semaphorin receptor-ligand pair of guidance molecules. The PlexinA4 receptor is expressed in the amacrine cells, whereas the transmembrane Semaphorin 6A ligand is concentrated in sublaminae S3-S5, thereby forming a repulsive barrier for PlexinA4-positive neurites (Fig. 17.10). Removal of either molecule results in neurite mistargeting.

These two examples of attractive and repulsive mechanisms for the laminar restriction of neurites and synapses explain a small number of the many selective interactions observed in the neuronal circuits of the retina. The intricate pattern of laminar restriction of the multitude of additional cellular interactions is likely achieved by a combination of cues that target neurites to or exclude them from specific domains. It is an open question whether synaptic specificity in neuronal networks, which are not organized in anatomical laminar structures, follows similar principles or whether new, yet unanticipated mechanisms are at work.

Summary

The diversity of synapses found in the nervous system is likely to be matched by an equally diverse array of mechanisms producing the synapses. These mechanisms are only now beginning to be elucidated. Prime candidates at present for synaptogenic roles include transmembrane proteins interacting with cognate receptors on the apposing cell surface (e.g., neuroligin/neurexin; RPTPs/TrkC receptors), diffusible molecules that instruct synaptic development in target cells (e.g., FGF22) or that interact directly with synaptic components (e.g., NARP, Cbln1), and intracellular components that act locally to organize synaptic constituents (e.g., gephyrin). Advances in visualizing neuronal connectivity in the intact brain have paved the way to the identification of signals that inhibit (Plexin-Semaphorin) or promote (cadherin-9) interactions between specific synaptic partners providing the first insights into synaptic specificity.

References

Aberle, H., Haghighi, A. P., Fetter, R. D., McCabe, B. D., Magalhaes, T. R., & Goodman, C. S. (2002). Wishful thinking encodes a BMP type II receptor that regulates synaptic growth in Drosophila. *Neuron, 33,* 545–558.

Arber, S., Burden, S. J., & Harris, A. J. (2002). Patterning of skeletal muscle. *Current Opinion in Neurobiology, 12,* 100–103.

Beeson, D., Higuchi, O., Palace, J., Cossins, J., Spearman, H., Maxwell, S., … Yamanashi, Y. (2006). Dok-7 mutations underlie a neuromuscular junction synaptopathy. *Science, 313,* 1975–1978.

Bergamin, E., Hallock, P. T., Burden, S. J., & Hubbard, S. R. (2010). The cytoplasmic adaptor protein Dok7 activates the receptor tyrosine kinase MuSK via dimerization. *Molecular Cell, 39,* 100–109.

Burden, S. J., Sargent, P. B., & McMahan, U. J. (1979). Acetylcholine receptors in regenerating muscle accumulate at original synaptic sites in the absence of the nerve. *The Journal of Cell Biology, 82,* 412–425.

Burgess, R. W., Skarnes, W. C., & Sanes, J. R. (2000). Agrin isoforms with distinct amino termini: differential expression, localization, and function. *The Journal of Cell Biology, 151,* 41–52.

Chih, B., Engelman, H., & Scheiffele, P. (2005). Control of excitatory and inhibitory synapse formation by neuroligins. *Science, 307,* 1324–1328.

Dalva, M. B., Takasu, M. A., Lin, M. Z., Shamah, S. M., Hu, L., Gale, N. W., & Greenberg, M. E. (2000). EphB receptors interact with NMDA receptors and regulate excitatory synapse formation. *Cell, 103,* 945–956.

DeChiara, T. M., Bowen, D. C., Valenzuela, D. M., Simmons, M. V., Poueymirou, W. T., Thomas, S., … Yancopoulos, G. D. (1996). The receptor tyrosine kinase MuSK is required for neuromuscular junction formation in vivo. *Cell, 85,* 501–512.

de Wit, J., Sylwestrak, E., O'Sullivan, M. L., Otto, S., Tiglio, K., Savas, J. N., … Ghosh, A. (2009). LRRTM2 interacts with Neurexin1 and regulates excitatory synapse formation. *Neuron, 64,* 799–806.

Engel, A. G., Ohno, K., & Sine, S. M. (2003). Sleuthing molecular targets for neurological diseases at the neuromuscular junction. *Nature reviews. Neuroscience, 4,* 339–352.

Eroglu, C., & Barres, B. A. (2010). Regulation of synaptic connectivity by glia. *Nature, 468,* 223–231.

Escher, P., Lacazette, E., Courtet, M., Blindenbacher, A., Landmann, L., Bezakova, G., … Brenner, H. R. (2005). Synapses form in skeletal muscles lacking neuregulin receptors. *Science, 308,* 1920–1923.

Flanagan-Steet, H., Fox, M. A., Meyer, D., & Sanes, J. R. (2005). Neuromuscular synapses can form *in vivo* by incorporation of initially aneural postsynaptic specializations. *Development, 132,* 4471–4481.

Fox, M. A., Sanes, J. R., Borza, D. B., Eswarakumar, V. P., Fassler, R., Hudson, B. G., … Umemori, H. (2007). Distinct target-derived signals organize formation, maturation, and maintenance of motor nerve terminals. *Cell, 129,* 179–193.

Francis, M. M., Evans, S. P., Jensen, M., Madsen, D. M., Mancuso, J., Norman, K. R., & Maricq, A. V. (2005). The Ror receptor tyrosine kinase CAM-1 is required for ACR-16-mediated synaptic transmission at the C. elegans neuromuscular junction. *Neuron, 46,* 581–594.

Gally, C., Eimer, S., Richmond, J. E., & Bessereau, J. L. (2004). A transmembrane protein required for acetylcholine receptor clustering in Caenorhabditis elegans. *Nature, 431,* 578–582.

Garner, C. C., Zhai, R. G., Gundelfinger, E. D., & Ziv, N. E. (2002). Molecular mechanisms of CNS synaptogenesis. *Trends in Neurosciences, 25,* 243–251.

Gautam, M., Noakes, P. G., Mudd, J., Nichol, M., Chu, G. C., Sanes, J. R., & Merlie, J. P. (1995). Failure of postsynaptic specialization to develop at neuromuscular junctions of rapsyn-deficient mice. *Nature, 377,* 232–236.

Glass, D. J., Bowen, D. C., Stitt, T. N., Radziejewski, C., Bruno, J., Ryan, T. E., … Yancopoulos, G. D. (1996). Agrin acts via a MuSK receptor complex. *Cell, 85,* 513–523.

Grady, R. M., Zhou, H., Cunningham, J. M., Henry, M. D., Campbell, K. P., & Sanes, J. R. (2000). Maturation and maintenance of the neuromuscular synapse: genetic evidence for roles of the dystrophin-glycoprotein complex. *Neuron, 25,* 279–293.

Graf, E. R., Zhang, X., Jin, S. X., Linhoff, M. W., & Craig, A. M. (2004). Neurexins induce differentiation of GABA and glutamate postsynaptic specializations via neuroligins. *Cell, 119,* 1013–1026.

Hall, A. C., Lucas, F. R., & Salinas, P. C. (2000). Axonal remodeling and synaptic differentiation in the cerebellum is regulated by WNT-7a signaling. *Cell, 100,* 525–535.

Hallock, P. T., Xu, C. F., Park, T. J., Neubert, T. A., Curran, T., & Burden, S. J. (2010). Dok-7 regulates neuromuscular synapse formation by recruiting Crk and Crk-L. *Genes & development, 24,* 2451–2461.

Herbst, R., Avetisova, E., & Burden, S. J. (2002). Restoration of synapse formation in Musk mutant mice expressing a Musk/Trk chimeric receptor. *Development, 129,* 5449–5460.

Hippenmeyer, S., Shneider, N. A., Birchmeier, C., Burden, S. J., Jessell, T. M., & Arber, S. (2002). A role for neuregulin1 signaling in muscle spindle differentiation. *Neuron, 36,* 1035–1049.

Jaworski, A., & Burden, S. J. (2006). Neuromuscular synapse formation in mice lacking motor neuron- and skeletal muscle-derived Neuregulin-1. *The Journal of Neuroscience: the Official Journal of the Society for Neuroscience, 26,* 655–661.

Jennings, C. G., Dyer, S. M., & Burden, S. J. (1993). Muscle-specific trk-related receptor with a kringle domain defines a distinct class of receptor tyrosine kinases. *Proceedings of the National Academy of Sciences of the United States of America, 90,* 2895–2899.

Jin, Y. (2002). Synaptogenesis: insights from worm and fly. *Current Opinion Neurobiology, 12,* 71–79.

Jing, L., Lefebvre, J. L., Gordon, L. R., & Granato, M. (2009). Wnt signals organize synaptic prepattern and axon guidance through the zebrafish unplugged/MuSK receptor. *Neuron, 61,* 721–733.

Kim, N., & Burden, S. J. (2008). MuSK controls where motor axons grow and form synapses. *Nature Neuroscience, 11,* 19–27.

Kim, E., & Sheng, M. (2004). PDZ domain proteins of synapses. *Nature Reviews. Neuroscience, 5,* 771–781.

Kim, N., Stiegler, A. L., Cameron, T. O., Hallock, P. T., Gomez, A. M., Huang, J. H., … Burden, S. J. (2008). Lrp4 is a receptor for Agrin and forms a complex with MuSK. *Cell, 135,* 334–342.

Kneussel, M., & Betz, H. (2000). Clustering of inhibitory neurotransmitter receptors at developing postsynaptic sites: the membrane activation model. *Trends in Neurosciences, 23,* 429–435.

Kummer, T. T., Misgeld, T., & Sanes, J. R. (2006). Assembly of the postsynaptic membrane at the neuromuscular junction: paradigm lost. *Current Opinion in Neurobiology, 16,* 74–82.

Lin, W., Burgess, R. W., Dominguez, B., Pfaff, S. L., Sanes, J. R., & Lee, K. F. (2001). Distinct roles of nerve and muscle in postsynaptic differentiation of the neuromuscular synapse. *Nature, 410,* 1057–1064.

Lin, W., Dominguez, B., Yang, J., Aryal, P., Brandon, E. P., Gage, F. H., & Lee, K. F. (2005). Neurotransmitter acetylcholine negatively regulates neuromuscular synapse formation by a Cdk5-dependent mechanism. *Neuron, 46,* 569–579.

Matsuoka, R. L., Nguyen-Ba-Charvet, K. T., Parray, A., Badea, T. C., Chedotal, A., & Kolodkin, A. L. (2011). Transmembrane semaphorin signalling controls laminar stratification in the mammalian retina. *Nature, 470,* 259–263.

Mejat, A., Ramond, F., Bassel-Duby, R., Khochbin, S., Olson, E. N., & Schaeffer, L. (2005). Histone deacetylase 9 couples neuronal activity to muscle chromatin acetylation and gene expression. *Nature Neuroscience, 8,* 313–321.

Misgeld, T., Kummer, T. T., Lichtman, J. W., & Sanes, J. R. (2005). Agrin promotes synaptic differentiation by counteracting an inhibitory effect of neurotransmitter. *Proceedings of the National Academy of Sciences of the United States of America, 102,* 11088–11093.

Missler, M., Zhang, W., Rohlmann, A., Kattenstroth, G., Hammer, R. E., Gottmann, K., & Sudhof, T. C. (2003). Alpha-neurexins couple $Ca2^{+}$ channels to synaptic vesicle exocytosis. *Nature, 423,* 939–948.

Nishimune, H., Sanes, J. R., & Carlson, S. S. (2004). A synaptic laminin-calcium channel interaction organizes active zones in motor nerve terminals. *Nature, 432,* 580–587.

O'Brien, R. J., Xu, D., Petralia, R. S., Steward, O., Huganir, R. L., & Worley, P. (1999). Synaptic clustering of AMPA receptors by the extracellular immediate-early gene product Narp. *Neuron, 23,* 309–323.

Okada, K., Inoue, A., Okada, M., Murata, Y., Kakuta, S., Jigami, T., … Yamanashi, Y. (2006). The muscle protein Dok-7 is essential for neuromuscular synaptogenesis. *Science, 312,* 1802–1805.

Packard, M., Koo, E. S., Gorczyca, M., Sharpe, J., Cumberledge, S., & Budnik, V. (2002). The Drosophila Wnt, wingless, provides an essential signal for pre- and postsynaptic differentiation. *Cell, 111,* 319–330.

Panzer, J. A., Song, Y., & Balice-Gordon, R. J. (2006). *In vivo* imaging of preferential motor axon outgrowth to and synaptogenesis at pre-patterned acetylcholine receptor clusters in embryonic zebrafish skeletal muscle. *The Journal of Neuroscience: the Official Journal of the Society for Neuroscience, 26,* 934–947.

Poulopoulos, A., Aramuni, G., Meyer, G., Soykan, T., Hoon, M., Papadopoulos, T., … Varoqueaux, F. (2009). Neuroligin 2 drives postsynaptic assembly at perisomatic inhibitory synapses through gephyrin and collybistin. *Neuron, 63,* 628–642.

Rapti, G., Richmond, J., & Bessereau, J. L. (2011). A single immunoglobulin-domain protein required for clustering acetylcholine receptors in C. elegans. *The EMBO Journal, 30,* 706–718.

Sanes, J. R., Marshall, L. M., & McMahan, U. J. (1978). Reinnervation of muscle fiber basal lamina after removal of myofibers. Differentiation of regenerating axons at original synaptic sites. *The Journal of Cell Biology, 78,* 176–198.

Scheiffele, P., Fan, J., Choih, J., Fetter, R., & Serafini, T. (2000). Neuroligin expressed in nonneuronal cells triggers presynaptic development in contacting axons. *Cell, 101,* 657–669.

Smith, C. L., Mittaud, P., Prescott, E. D., Fuhrer, C., & Burden, S. J. (2001). Src, Fyn, and Yes are not required for neuromuscular synapse formation but are necessary for stabilization of agrin-induced clusters of acetylcholine receptors. *The Journal of Neuroscience: the Official Journal of the Society for Neuroscience, 21,* 3151–3160.

Son, Y. J., & Thompson, W. J. (1995). Schwann cell processes guide regeneration of peripheral axons. *Neuron, 14,* 125–132.

Sudhof, T. C. (2008). Neuroligins and neurexins link synaptic function to cognitive disease. *Nature, 455,* 903–911.

Takahashi, H., Arstikaitis, P., Prasad, T., Bartlett, T. E., Wang, Y. T., Murphy, T. H., & Craig, A. M. (2011). Postsynaptic TrkC and presynaptic PTPsigma function as a bidirectional excitatory synaptic organizing complex. *Neuron, 69,* 287–303.

Tang, H., & Goldman, D. (2006). Activity-dependent gene regulation in skeletal muscle is mediated by a histone deacetylase (HDAC)-Dach2-myogenin signal transduction cascade. *Proceedings of the National Academy of Sciences of the United States of America, 103,* 16977–16982.

Uemura, T., Lee, S. J., Yasumura, M., Takeuchi, T., Yoshida, T., Ra, M., … Mishina, M. (2010). Trans-synaptic interaction of GluRdelta2 and Neurexin through Cbln1 mediates synapse formation in the cerebellum. *Cell, 141,* 1068–1079.

Umemori, H., Linhoff, M. W., Ornitz, D. M., & Sanes, J. R. (2004). FGF22 and its close relatives are presynaptic organizing molecules in the mammalian brain. *Cell, 118,* 257–270.

Valenzuela, D. M., Stitt, T. N., DiStefano, P. S., Rojas, E., Mattsson, K., Compton, D. L., … Yancopoulos, G. D. (1995). Receptor tyrosine kinase specific for the skeletal muscle lineage: expression in embryonic muscle, at the neuromuscular junction, and after injury. *Neuron, 15,* 573–584.

Varoqueaux, F., Aramuni, G., Rawson, R. L., Mohrmann, R., Missler, M., Gottmann, K., … Brose, N. (2006). Neuroligins determine synapse maturation and function. *Neuron, 51,* 741–754.

Weatherbee, S. D., Anderson, K. V., & Niswander, L. A. (2006). LDL-receptor-related protein 4 is crucial for formation of the neuromuscular junction. *Development, 133,* 4993–5000.

Weston, C., Gordon, C., Teressa, G., Hod, E., Ren, X. D., & Prives, J. (2003). Cooperative regulation by Rac and Rho of agrin-induced acetylcholine receptor clustering in muscle cells. *The Journal of Biological Chemistry, 278,* 6450–6455.

Weston, C., Yee, B., Hod, E., & Prives, J. (2000). Agrin-induced acetylcholine receptor clustering is mediated by the small guanosine triphosphatases Rac and Cdc42. *The Journal of Cell Biology, 150,* 205–212.

Williams, M. E., de Wit, J., & Ghosh, A. (2010). Molecular mechanisms of synaptic specificity in developing neural circuits. *Neuron, 68,* 9–18.

Williams, M. E., Wilke, S. A., Daggett, A., Davis, E., Otto, S., Ravi, D., … Ghosh, A. (2011). Cadherin-9 regulates synapse-specific differentiation in the developing hippocampus. *Neuron, 71,* 640–655.

Yamagata, M., & Sanes, J. R. (2008). Dscam and Sidekick proteins direct lamina-specific synaptic connections in vertebrate retina. *Nature, 451,* 465–469.

Yang, X., Arber, S., William, C., Li, L., Tanabe, Y., Jessell, T. M., … Burden, S. J. (2001). Patterning of muscle acetylcholine receptor gene expression in the absence of motor innervation. *Neuron, 30,* 399–410.

Yang, X., Li, W., Prescott, E. D., Burden, S. J., & Wang, J. C. (2000). DNA topoisomerase IIbeta and neural development. *Science, 287,* 131–134.

Zhang, B., Luo, S., Wang, Q., Suzuki, T., Xiong, W. C., & Mei, L. (2008). LRP4 serves as a coreceptor of agrin. *Neuron, 60,* 285–297.

Steven J. Burden and Peter Scheiffele

This page intentionally left blank

CHAPTER

18

Programmed Cell Death and Neurotrophic Factors

Major hallmarks of embryonic development include the generation of new cells, the growth of tissues and organs, and the gradual but inexorable acquisition of new cellular and morphological properties (phenotypes). Accordingly, in the past, a major focus of developmental neurobiologists has been the study of progressive events, such as those described in several other chapters in this section. In this context, the concept of significant regressive events, such as programmed cell death (PCD), occurring during development, initially was considered counterintuitive (Oppenheim, 1981; 1991). Subsequently, however, it has been demonstrated that cell loss and other regressive events, including synapse elimination (Chapter 19), are the rule rather than the exception. In most developing tissues that have been examined in multicellular organisms, substantial cell loss occurs. As we discuss later, even a significant proportion of adult-generated neurons in the hippocampus and olfactory system also undergo normal PCD (Gage, Kempermann, & Song, 2008). Genetic programs that result in cell death have been suggested to be a default pathway for all cells. In this view neurons would escape this default fate only by receiving the appropriate receptor-mediated survival signals (e.g., neurotrophic factors, NTFs) that induce changes in the molecular and biochemical events required for cell survival and neuronal differentiation. In the absence of these ligands, the unbound receptors may activate death pathways and have been called "dependence-receptors." Neurons that express dependence receptors require—that is, are dependent upon—the presence of ligands for their anti-apoptotic role in promoting survival; however, in the absence of ligands, these neurons do not die by default, but rather the unbound receptors are *actively* pro-apoptotic and trigger cell death. Aberrations in the mechanisms regulating the balance between cell proliferation and cell elimination can lead to pathologies of tissue growth and differentiation by reduced cell death (e.g., cancer) or excessive neuronal death (e.g., neurodegenerative disease). Mechanisms that regulate the balance between cell production and death serve to determine the ultimate number and maintenance of neurons in the nervous system. Genes regulating these mechanisms provide an underlying substrate for evolutionary changes in brain size and structure and ultimately in adaptive behavior.

Both progressive and regressive events during development are regulated by intercellular signals. Somewhat surprisingly, many of the same intercellular signals (e.g., NTFs) that contribute to the regulation of progressive events during nervous system development also contribute to the control of regressive events. Neurotrophic factors are now appreciated for their role in both survival- and nonsurvival-related activities in both developing and mature nervous systems where they play important roles in the regulation of activity-dependent structural and functional plasticity of the nervous system. Accordingly, we include here a review of both survival and nonsurvival functions of NTFs.

Because neurons may die for a variety of reasons and in many different situations, it is important to describe the type of cell death observed most often in the developing nervous system. Although the loss of cells during normal development has been called by many different names (normal cell death, spontaneous cell death, naturally occurring cell death, apoptosis, and developmental cell death), we prefer the term *programmed cell death*

© 2013 Elsevier Inc. All rights reserved.

(PCD). PCD is defined as the spatially and temporally reproducible and species-specific loss of large numbers of individual cells both during development and, in many tissues and organs, throughout life, as adult cells die and are replaced by precursor or stem cells as a means of maintaining stable structure and function (homeostasis). Accidental, injury-induced, pathological, and disease-related forms of cell death are not included even though we recognize that the biochemical and molecular mechanisms used to kill cells in these situations may overlap with those involved in developmental PCD and in some cases, like PCD, the death may even be genetically controlled. This definition of PCD also makes no *a priori* assumptions about either the morphological or the biochemical pathways by which cells die or the stimuli that trigger cell death. The use of the word "programmed" refers to the reproducible, spatiotemporally species-specific occurrence of cell loss and is not meant to imply that the cell loss is genetically predetermined, inherited from precursor cells, or inevitable. Finally, and as we discuss in more detail later, PCD is not synonymous with the term *apoptosis*, which refers only to one specific, albeit common, mode of developmental cell death.

CELL DEATH AND THE NEUROTROPHIC HYPOTHESIS

Early embryological studies of the interactions between developing neurons and their peripheral targets laid the foundation for the discovery of cell death and the first NTF (Cowan, 2001). These studies formed the conceptual framework for the neurotrophic ("nerve feeding") hypothesis (Fig. 18.1). By 1949, Viktor Hamburger and Rita Levi-Montalcini had provided compelling evidence to support the significance of normal embryonic neuronal death and postulated that target-derived signals act to regulate the number of neurons that survive embryonic development. In subsequent studies, they and their colleagues identified a specific protein, nerve growth factor (NGF), that influenced development of the same populations of neurons (sensory and sympathetic) that their earlier studies had suggested were regulated by target-derived signals.

By 1960, the preparation of specific antibodies that block NGF activity allowed Stanley Cohen, Levi-Montalcini, and co-investigators to demonstrate the almost total degeneration of sympathetic ganglia *in vivo* following the specific deprivation of NGF activity. Only with the evidence from this "immunosympathectomy" (the deletion of the sympathetic system by antibody treatment) was NGF considered to be an endogenous survival or maintenance factor for these neurons. Three decades after the original discovery that an unknown chemical substance produced by tumor cells affects the development of sensory and sympathetic neurons, it was finally recognized that NGF was the hypothetical target-derived trophic signal first postulated by Hamburger and Levi-Montalcini in 1949 to be involved in regulating the number of surviving neurons in these populations during normal development. Historically, then, the appreciation of the role of PCD in the developing nervous system and the discovery of the first target-derived NTF (NGF) were inextricably linked; together they provided the foundation for the neurotrophic hypothesis, and both owe their origin to earlier studies of interactions between developing neurons and their synaptic targets (Oppenheim, 1991).

Despite the significance of this conceptual advance and the singular role of the neurotrophic hypothesis in guiding research in this field for over half a century, new evidence indicates that the neurotrophic hypothesis is in need of revision in order to encompass the increased appreciation that the cell–cell interactions that regulate neuronal survival and death are more complex than originally thought. It is now clear, for example, that the efferent targets of neurons are only one of several sources of survival-promoting NTFs and that NGF is only one of many diverse NTFs that regulate development and survival.

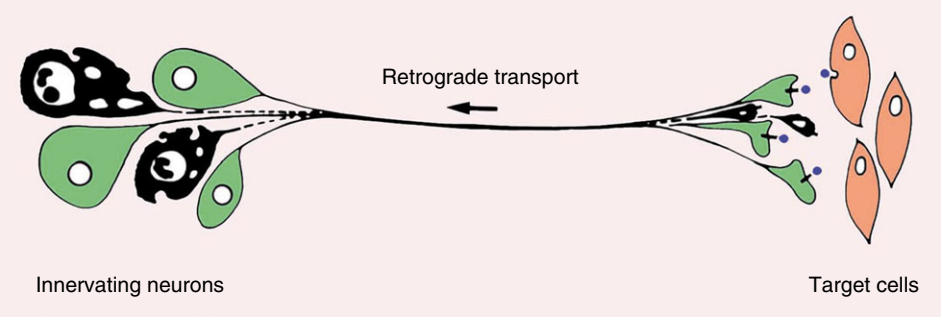


FIGURE 18.1 The neurotrophic hypothesis postulates that developing neurons survive (green neurons) only when they successfully compete for target-derived trophic molecules (blue) that are internalized at the nerve terminal and are transported retrogradely along the axon (arrow) to the cell body. Neurons that fail to obtain a sufficient flow of trophic molecules from the target die by programmed cell death (black neurons with condensed chromatin). *Modified and reproduced with permission from Barde (1989).*

THE ORIGINS OF PROGRAMMED CELL DEATH AND ITS WIDESPREAD OCCURRENCE IN THE DEVELOPING NERVOUS SYSTEM

PCD has been identified in plants, several species of unicellular eukaryotes, including yeast, as well as in prokaryotes such as bacteria, one of the oldest forms of life on earth (Ameisen, 2004). One likely explanation for the evolutionary origins of PCD in eukaryotes was the appropriation of cell death–associated genes from bacterial mitochondria by a process of endosymbiosis. Based on the principle of genetic pleiotropy, it seems likely that under the right conditions genes required for vital functions such as metabolism, proliferation, and differentiation can also serve pro- and anti-cell death functions and that this pleiotropy played a major role in their later selection in unicellular eukaryotes and metazoans. Even some of the core genetic machinery currently employed for regulating PCD in mammals, including neuronal PCD, has retained a degree of pleiotropy. Both caspases and Bcl-2 family members mediate a number of nondeath functions in both neuronal and nonneuronal cells (Jiao & Li, 2011).

After being incorporated into the eukaryotic genome, this core cell death machinery expanded in complexity over time to include the diversity found in extant multicellular organisms. Once the pro- and anti-apoptotic machinery was in place in the genome of multicellular animals, it could then be co-opted to mediate the many diverse roles now served by PCD, including different functions of PCD in developing neurons. Finally, a word of explanation is needed regarding the seemingly counterintuitive existence of PCD in unicellular organisms—that is, how can the death of individual organisms be beneficial? In fact, several adaptive roles of PCD in these organisms have been reported (Nedelcu, Driscoll, Durand, Herron, & Raskidi, 2010).

Although a systematic taxonomic survey of all representative species with a nervous system has not been done, nonetheless the available evidence is consistent with the idea that some PCD of developing neurons occurs in all such organisms. For many invertebrate and vertebrate species, the PCD of neurons involves virtually all regions and cell types in the central and peripheral nervous system. Motoneurons, sensory neurons, autonomic neurons, and both long projection and local circuit interneurons in the brain and spinal cord all undergo restricted periods of PCD. Although the magnitude of neuronal cell death varies from population to population, as many as one-half or more of all cells in a population will die during development (Fig. 18.2). In some special cases, such as the loss of transient neuronal structures during insect and amphibian metamorphosis (e.g., the Rohon-Beard sensory neurons in fish and frogs), most or all cells die. Cell death in the nervous system thus clearly occurs on a very large scale, indicating that it plays fundamental and essential roles in normal development.

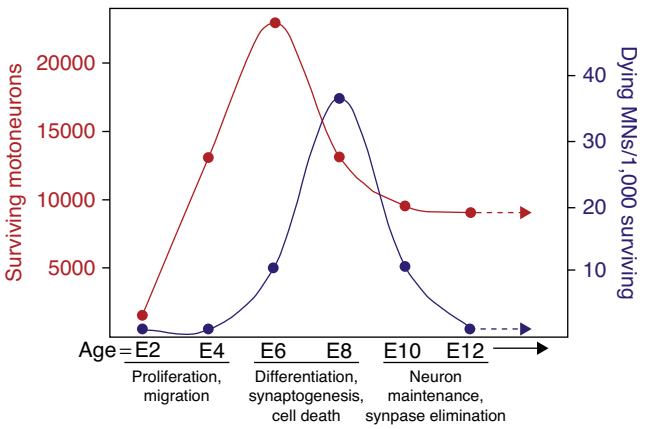


FIGURE 18.2 Changes in the number of developing lumbar spinal motoneurons in the chick embryo (in red) compared to the number of dying motoneurons (in blue) during the first half of the 21 day incubation period. Approximately 50 to 60% of these motoneurons undergo PCD between embryonic day (E)6 and E12.

Studies of cell death in the nervous system have focused on the loss of developing postmitotic neurons as they form synaptic connections with targets and afferents. However, extensive PCD also occurs during neurulation and in mitotically active neural progenitor cells, as well as in postmitotic but undifferentiated neurons in the early neural tube (Fig. 18.3). Accordingly, PCD in the nervous system is not limited to any particular stage of development. The loss of cells at different developmental stages probably serves distinct functions and may be mediated by different mechanisms. PCD also occurs in central and peripheral glial cells. For example, like neurons, myelin-forming oligodendrocytes in the optic nerve and Schwann cells in peripheral nerves are overproduced, and many die by PCD. Their loss is thought to reflect a competition for axon-derived trophic signals, the end result of which is the survival of an appropriate number of glial cells for optimal myelination of the available axons (Fig. 18.4).

Because PCD is the normal differentiated or terminal fate of many developing cells, commitment to this fate occurs in much the same way as the phenotypic fate of cells destined to survive in the embryo. Developmental biologists have identified two major ways in which the commitment of a cell to a particular differentiated phenotype occurs. The first mechanism, intrinsic or *autonomous specification*, involves the segregation of critical cytoplasmic molecules by asymmetric cell division; cell death thus is programmed into the lineages that generate somatic cells. The second mechanism of commitment involves extrinsic signals from other cells and is

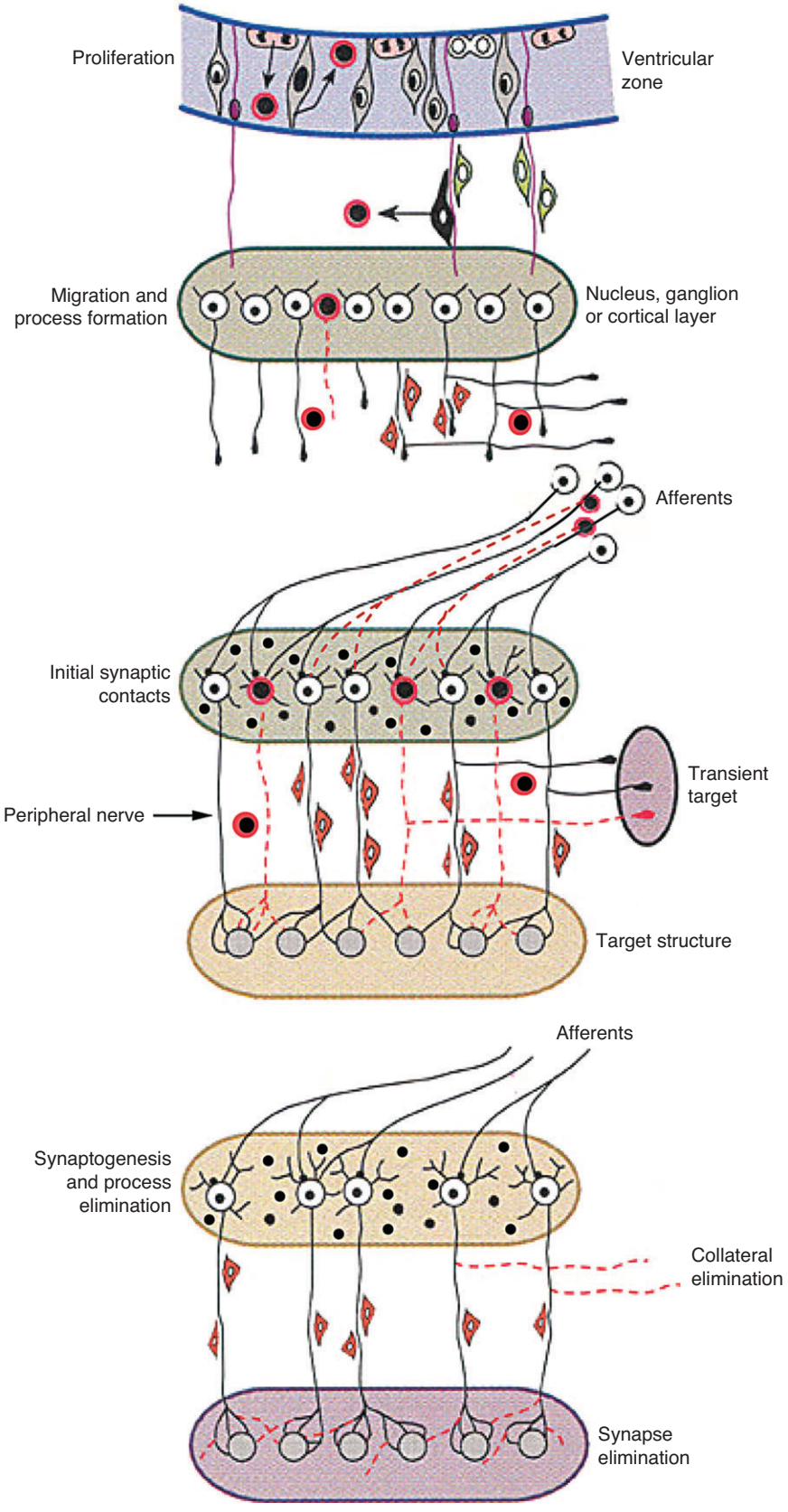


FIGURE 18.3 Schematic illustration of some key steps in neuronal development. Neurons undergoing PCD (●) are observed during neurogenesis in the ventricular zone, during migration and while establishing synaptic contacts. Schwann cells in developing nerves also undergo PCD. ◗ Represents peripheral glial (Schwann) cells; ● represents CNS glia (astrocytes, oligodendrocytes); ⊙ represents surviving, differentiating neurons (e.g., motoneurons whose targets are skeletal muscle and other neurons in the CNS, with neuronal targets).

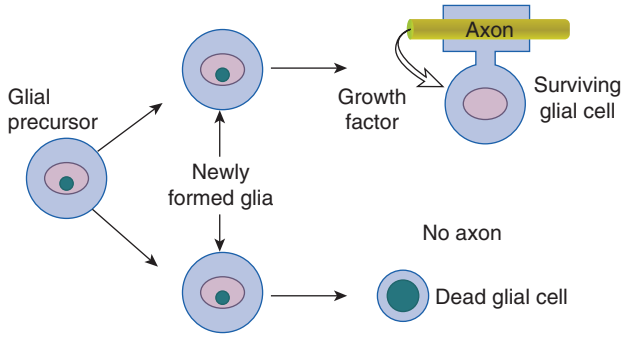


FIGURE 18.4 The PCD of glial cells is regulated by axon-derived signals. Glial cells that fail to compete successfully for these signals undergo PCD.

called *conditional specification*; initially, the cells have the potential to follow more than one path of differentiation, but as development proceeds, signals from other cells act to gradually limit and specify cell fate. Although all organisms use a combination of autonomous and conditional developmental strategies, as a general rule invertebrates are more likely to utilize autonomous specification, whereas most vertebrates exhibit conditional specification. Despite this distinction, as described later, many of the genetic and molecular pathways for PCD are remarkably similar in invertebrates and vertebrates. In addition to these two major pathways for fate determination, as we discuss below, there is increasing evidence for a third pathway involving stochasticity: random fluctuations (noise) in gene expression (fate determination by the "roll of the dice").

Summary

The PCD of developing neurons appears to occur in virtually all vertebrate and invertebrate species. More generally, PCD also occurs in many different cells and tissues of unicellular and multicellular organisms and thus may have arisen very early during evolution. With few exceptions, PCD occurs in virtually all types of developing neurons and can take place at stages of development from the time of proliferation until the establishment of synaptic connections. Developing glial cells also exhibit PCD. Cell death is the terminal phenotypic fate of subpopulations of developing neuronal and glial cells and, like other cell fate decisions, is controlled by both intrinsic and extrinsic signals.

FUNCTIONS OF NEURONAL PROGRAMMED CELL DEATH

Why does PCD occur? This is a reasonable question to ask because the loss of large numbers of developing neurons is counterintuitive. Why should embryos invest precious resources in generating cells and tissues only to later cast many of these aside? A satisfactory answer to this apparent paradox requires an evolutionary perspective that addresses two central aspects of the problem. First, how did the biochemical machinery (the cell death program) needed to actively kill cells arise? Second, why, in many developing tissues, are more cells generated than are apparently needed?

The first question was addressed in the previous section; however, understanding the evolution of the biochemical cell death program does not help answer the second question of why there is often a massive overproduction of neurons during development that are later eliminated by PCD. Two explanations have been offered (Oppenheim, 1991). First, each case of PCD may have evolved to serve a distinct biological function. For example, according to this view, in the case of spinal motoneurons, natural selection is thought to be directly responsible for both the overproduction and the subsequent death of neurons as a means for creating an optimal (adaptive) level of functional muscle innervation (e.g., systems-matching).

The second view is that the overproduction of neurons (or other cells) is an inevitable outcome of the imprecise kinetics of proliferation of precursor cells. Once too many (or too few) cells are generated and the pro- and antiapoptotic machinery is in place, natural selection then acts via regulated cell survival and death to mediate a variety of different adaptive needs. For example, following the loss or absence of essential survival signals, the death of excess cells could be accomplished easily by co-opting the cell death machinery that evolved to kill cells in early eukaryotes. In reality, both of the proposed mechanisms may occur (Table 18.1). Many of the circumstances in which the PCD of neurons occurs are thought to mediate distinct adaptive functions, as summarized in Table 18.1. In many of these examples, the production of excess neurons provides a substrate on which differential survival or PCD can then act to meet a variety of putative adaptive needs. Similarly, if fewer numbers of cells are generated, then compensation, by reducing the extent of cell death, provides a means for ensuring, for example, that a sufficient number of neurons are still available for optimal target innervation. As new genes are identified that regulate PCD, and as we gain a better understanding of how cellular and molecular signals control cell death and survival, there will be increased opportunities for preventing PCD in select neuronal populations *in vivo* and directly assessing whether its occurrence is, in fact, adaptive.

Summary

The biochemical and molecular pathways that regulate cell death and survival arose early in the evolution of animal life. Once this cellular capacity arose, however, it is likely that it was co-opted to serve a variety of biological

TABLE 18.1 Some Possible Functions of Developmental PCD in the Nervous System[a]

1. Differential removal of cells in males and females (sexually dimorphic spinal motor nucleus in many mammals).
2. Deletion of some of the progeny of a specific sublineage that are not needed (loss of specific progeny of the AB blastomere that is involved in generating ring ganglia in *C. elegans*).
3. Negative selection of cells of an inappropriate phenotype (ligand/receptor-induced cell death in the early chick embryo retina?).
4. Pattern formation and morphogenesis (neurulation/neural tube closure; differential thickness of cortical layers).
5. Deletion of cells that act as transient targets or that provide transient guidance cues for axon projections (death of pioneer neurons/glia in insects).
6. Removal of cells and tissues that serve a transient physiological or behavioral function (loss of Rohon-Beard sensory neurons during metamorphosis in frogs).
7. Systems-matching by creating optimal quantitative innervation between interconnected groups of neurons and between neurons and their targets (see text).
8. Systems-matching between neurons and their glial partners by regulated glial PCD (Schwann cells and peripheral axons; see text).
9. Error correction by the removal of ectopically positioned neurons or of neurons with misguided axons or inappropriate synaptic connections (loss of ipsilaterally projecting retinal ganglion cells).
10. Removal of damaged or harmful cells (death of cells with DNA damage).
11. Regulation of the size of mitotically active neural progenitor populations (see text).
12. The production of excess neurons may serve as an ontogenetic buffer for accommodating mutations that require changes in neuronal numbers in order to be evolutionary adaptive (evolutionary increases in limb size may require increased sensory and motoneuron survival).
13. Regulated survival of subpopulations of adult-generated neurons as a means of experience-dependent plasticity (see text).

[a]*For references, see Buss, Sun, and Oppenheim (2006).*

functions. In the nervous system, some major functions include establishing optimal levels of connectivity between neuronal populations, eliminating aberrant cells or connections, regulating the size of progenitor populations, and serving transient functional or other needs of immature animals.

MODES OF CELL DEATH IN DEVELOPING NEURONS

The specific morphological appearance exhibited by degenerating neurons can provide insight into the cellular and molecular mechanisms by which the cells are destroyed. Historically, pathologists were the first to be interested in this issue, and they focused on distinguishing different kinds of cell and tissue degeneration following disease, injury, and trauma (Clarke & Clarke, 1996). Over 125 years ago, the term *necrosis* was coined to describe what today comprises the major form of accidental or pathological degeneration. The pathological necrotic death of neurons following injury usually involves the degeneration of groups of contiguous cells in a region that initiates an inflammatory response that can be discerned easily in tissue sections (Fig. 18.5). At about the same time that necrosis was first described, however, another form of cell degeneration, called spontaneous cell death, was observed in normal adult mammals. Spontaneous cell death (now called PCD) was thought to provide a means for counterbalancing mitosis in adult tissues in which the turnover of cells normally occurs. PCD typically involves the sporadic loss of individual cells in a population that often degenerate by a different mode from necrosis, one that does not involve inflammation. Early steps in the spontaneous PCD cascade leading up to when histological signs of frank degeneration first occur may take many hours or days. Once that point is reached, however, the degenerative process is rapid, with individual cells dying and being removed in minutes or a few hours. The loss of thousands of neurons over several days is the consequence of many rapid individual cell deaths that at any moment in time represent only a small minority (~1%) of all the cells in the population. For this reason, historically, the occurrence and magnitude of even massive PCD can (and often did) go unnoticed.

Two different strategies can be used to quantify neuron death: One can determine the total number of neurons at different stages and quantify the decrease (Fig. 18.2), or one can determine an increase in the number of dying (pyknotic) cells during the period of PCD and determine the timing and (with less precision) the extent of cell death (Fig. 18.5). The counting of neurons originally was done by counting profiles in thin sections and application of correction factors to account for multiple profiles per particle; a particle is defined as an identifiable object such as a cell, nucleus, or nucleolus (Clarke & Oppenheim, 1995); such counting is now increasingly replaced by stereological methods that count particles in thicker sections, a method that is less affected by confounding differences in particle size but that still presents challenges due to other sources of potential errors if not executed properly (Schmitz & Hof, 2005).

Cell degeneration in adult and developing tissues has often been dichotomized into death by either apoptosis or necrosis, a distinction based initially on morphological differences (Fig. 18.6). *Apoptosis* is a Greek word indicating the seasonal piecemeal dropping of leaves from a tree and originally was coined to describe all forms of cell death that

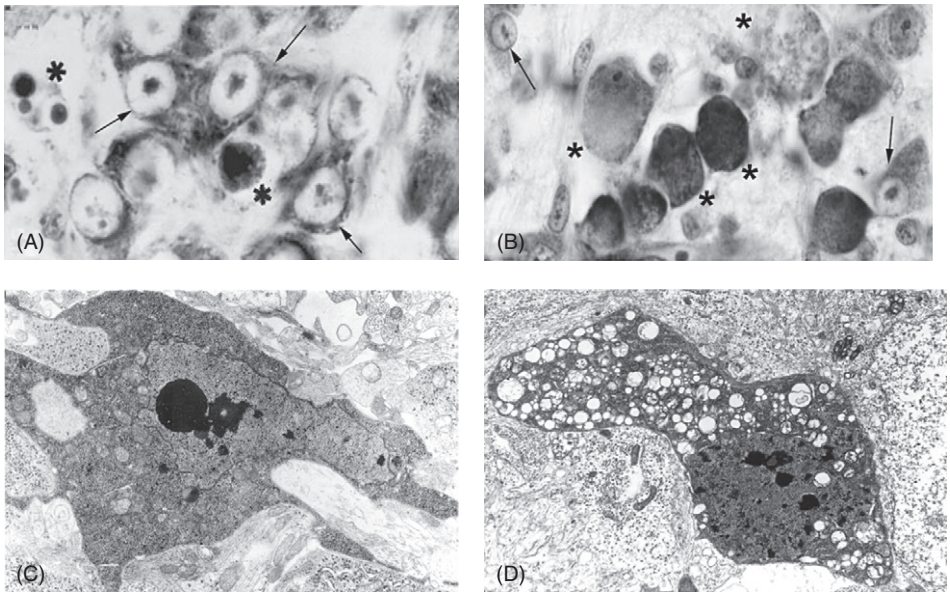


FIGURE 18.5 Spinal motoneurons in the chick embryo. (A) Ventral horn from a control embryo. Note that only two cells (asterisks) are undergoing apoptotic PCD; the others appear normal (arrows). (B) Ventral horn from an embryo following treatment with an excitotoxin. Most neurons are undergoing a necrotic type of cell death (asterisks) whereas some appear normal (arrows). Apoptotic (C) and necrotic motoneurons (D) in the chick embryo spinal cord as seen with an electron microscope.

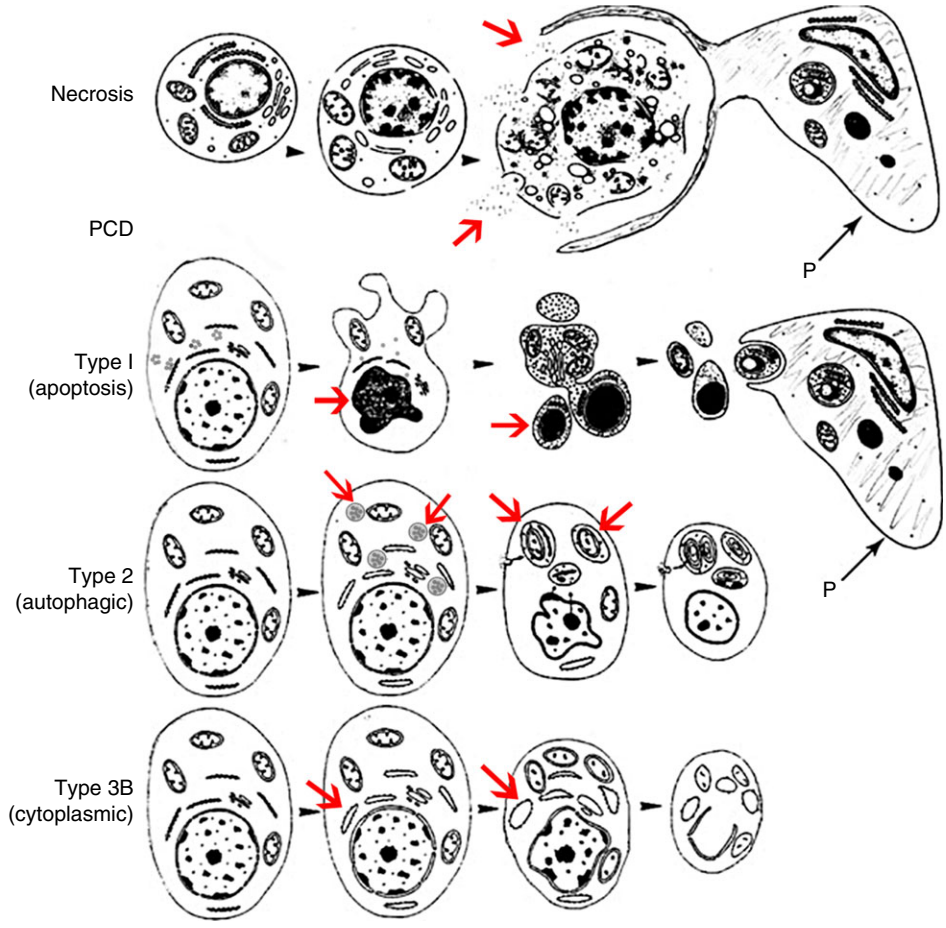


FIGURE 18.6 Schematic representation of cellular changes during necrotic cell death and during three of the most common types of PCD observed at the ultrastructural level. Key features of each type of cell death are indicated by the red arrows. Only type 1 PCD meets most of the criteria for defining apoptosis. The cells on the right marked P represent phagocytic cells engulfing necrotic cell corpses and apoptotic bodies. Phagocytosis also occurs in the other types of PCD but is not shown.

share certain morphological characteristics (Wyllie, Kerr, & Currie, 1980). Cells dying by apoptosis shrink in size, and the nuclear chromatin condenses and becomes pyknotic, whereas the cell membrane and cytoplasmic organelles tend to remain relatively intact. Eventually the cytoplasm and nucleus break up into membrane-bound apoptotic bodies that are phagocytized either by professional phagocytes (macrophages) or by healthy adjacent cells (e.g., glia

that serve as transient phagocytes). Phagocytes recognize dying cells by their expression of death-related cell surface signals. In contrast, necrosis involves an initial swelling of the cell, only modest condensation of chromatin, cytoplasmic vacuolization, breakdown of organelles, and rupture of the cell membrane allowing the release of cellular contents (causing inflammation), followed by shrinkage and loss of nuclear chromatin. Because necrotic cell death elicits an inflammatory response in adjacent tissue, macrophages derived from the immune system attack and phagocytize cellular debris. In contrast, cell death by apoptosis usually involves individual cells that are engulfed and phagocytized before they can release their cellular contents or induce an inflammatory response in adjacent tissue.

Another feature that has been used to distinguish between apoptotic and necrotic cell death is the occurrence during apoptosis of a specific form of chromosomal DNA fragmentation and degradation that is mediated by DNA-specific proteases. DNA digestion occurs at internucleosomal sites, producing small, double-stranded fragments of DNA that migrate in a ladder pattern in multiples of 180–200 bp after electrophoresis in agarose gels. This form of DNA fragmentation can also be visualized in tissue sections by a technique that labels the double-stranded DNA breaks associated with apoptosis. DNA fragmentation suggests that apoptosis is an active process undertaken by the cell, whereas necrosis is a passive process resulting from extracellular damage. However, a variety of toxic and traumatic stimuli, previously thought to induce necrosis, can also induce morphological signs of apoptosis and may be associated with changes in PCD- or apoptosis-associated genes (Bredesen, Rao, & Mehlen, 2006).

Within the context of the developing nervous system, the dichotomy of apoptosis versus necrosis is an oversimplification. Developing cells undergoing normal PCD as well as neurons dying following injury should be categorized by operational definitions that use morphological, genetic, and biochemical criteria. Studies of the developing nervous system at the ultrastructural level provided an initial characterization of neuronal death noting that there were features of dying cells that appeared to indicate that all neurons do not die by identical mechanisms (Clarke & Clarke, 1996). "Nuclear" (Type I) neuronal death featured a dissociation of polyribosomes to free ribosomes, Golgi fragmentation to numerous vesicles, swollen mitochondria, and condensation of nuclear chromatin; the cell takes on a reduced size and crenulated appearance, neurotubules and neurofilaments are grouped together, and lysosomes appear normal. This type of neuronal degeneration is consistent with apoptosis.

There are other types of neuronal death during development, however, that do not fall under the categories of apoptosis or necrosis. In "cytoplasmic" (Type III) degeneration, the cisternae of Golgi, nuclear envelope, and rough endoplasmic reticulum are swollen, together with a vacuolization of mitochondria. Polyribosomes appear normal, and while the nuclear chromatin shows some condensation, the nucleus appears relatively normal. There is another type of cell death referred to as autophagy or Type II cell death that is characterized by the presence of abundant lysosomes and autophagic vacuoles. The specific morphological events that a neuron undergoes during PCD may depend on the type of neuron, its state of differentiation, and the stimuli that induce cell death. For example, ciliary ganglion cells predominantly exhibit features of cytoplasmic cell death during PCD; however, when peripherally deprived of targets, the majority of ciliary cells exhibit features of nuclear or apoptotic cell death (Oppenheim, 1981). The PCD of developing vertebrate neurons provides a striking example of the problems encountered in attempting to rigidly classify the pathway of degeneration as being either necrotic or apoptotic. Importantly, evaluation of the morphological features of PCD can further the understanding of the biochemical pathways of PCD and the roles of the specific genes involved.

Summary

Degenerating cells are often categorized into two classes: death by apoptosis or death by necrosis. However, the occurrence of certain features of apoptosis in neurons following injury and the occurrence of several types of PCD other than apoptosis indicate that a more accurate means of identifying and defining distinct forms of cell death should now be employed.

THE MODE OF NEURONAL CELL DEATH REFLECTS THE ACTIVATION OF DISTINCT BIOCHEMICAL AND MOLECULAR MECHANISMS

As described previously, the normal death of cells in the developing nervous system has long been thought to be regulated by competition for NTFs (the Neurotrophic Hypothesis). Investigators thought that the doomed neurons, lacking sufficient amounts of an NTF to sustain normal metabolic events, passively degenerated by a process analogous to starvation. Beginning in the late 1980s, several lines of evidence forced a reappraisal of the view that cell death in the nervous system is a passive process and led to the demonstration that for many types of neurons, PCD is regulated by the interaction of specific genetic programs that either inhibit or induce degeneration.

PCD of some nonneuronal cells was known previously to be a metabolically active, ATP-dependent process. In the nervous system, inhibition of RNA or protein synthesis could also prevent the death of sympathetic ganglion cells deprived of NGF *in vitro* and motoneuron and

sensory neuron PCD *in vivo*. Additionally, in the early 1980s, genetic mutations in the nematode worm *Caenorhabditis elegans* that prevent PCD had been described, thereby providing further evidence that neuronal death is a genetically regulated, metabolically active process. Considerable progress has been made since the early 1990s in identifying cell death–associated genes and their pathways of action. Due to historical precedent, these have been classified as pro- and anti-apoptotic genes; however, it is important to point out that in many instances the genes involved may actually induce cell death with morphological features of both apoptotic and nonapoptotic degeneration (see the previous section).

Genetic studies of cell death in *C. elegans* first demonstrated that PCD is regulated by specific genes. Because the genetics, morphology, and cell lineages in *C. elegans* have been so well defined, this organism provides a particularly informative and powerful model for analyzing the molecular genetics of PCD (Horvitz, 2003). Of the approximately 1000 somatic cells generated (of which 302 are neurons and 56 are glial cells), 131 undergo embryonic PCD, and most of these are mitotically active neuron precursors. In each individual, the same cells die at specific times in development, and these corpses are then engulfed and degraded by neighboring cells. Despite the enormous evolutionary gap that separates the appearance of worms, flies, and vertebrates, significant homology exists in the structure and function of specific cell death pathways between these taxonomic groups.

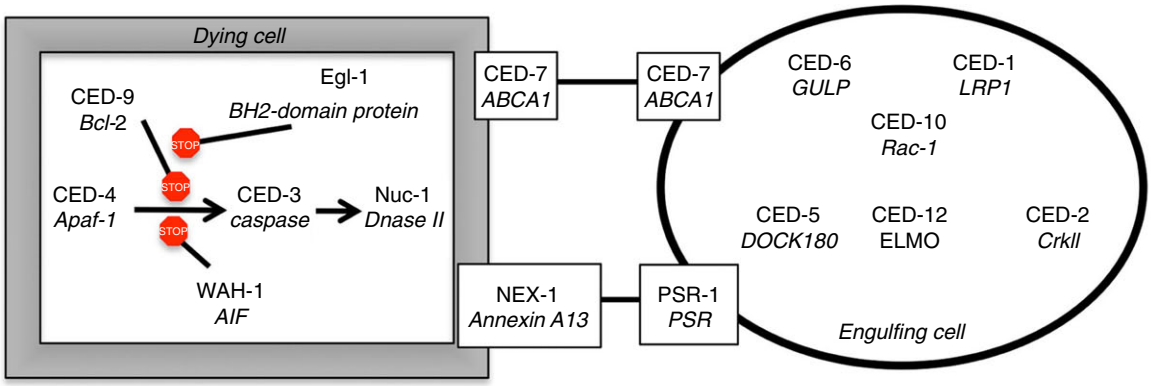
As summarized in Figure 18.7, there is a sequential cascade of genetically regulated steps involved in PCD. Upstream of the actual execution of the death process, transcription factors specify certain cell types for death, while sparing others. In *C. elegans*, manifestation of the death fate requires expression of two pro-apoptotic genes, ced–3 and ced–4 (ced, cell death), which together mediate the actual breakdown of cellular constituents. Prevention of cell death induced by ced-3 and ced-4 can occur by activation of the anti-apoptotic gene ced-9. Separate genes, such as nuclease-1 (nuc-1), are required for the degradation of DNA in dying cells, and several additional genes are involved in the recognition, engulfment, and processing of dead cellular corpses by phagocytes. Evidence for the involvement of this genetic cascade in cell death and survival in *C. elegans* comes from several different approaches, most notably from genetic studies of loss-of-function mutants. For example, in the absence of ced-9, many cells that normally survive die, whereas in the absence of ced-3 or ced-4, all PCD is prevented.

Identification of the DNA sequences of the major cell death genes ced-3, ced-4, and ced-9 in the worm in the 1980s and early 1990s resulted in the subsequent discovery of vertebrate and insect homologues that serve similar functions (Fig. 18.7). In mammals, ced-3 is represented by a large family of related proteases called caspases. Caspases are cysteine proteases that cleave after aspartate residues, and 12 distinct caspases have been cloned in human. In cell death, initiator pro-caspases are autolytically cleaved, and they then cleave effector pro-caspases. The effector caspases act on specific substrates that result in cell demise. Ced-4 is represented by a single vertebrate homologue, Apaf-1 (apoptosis protease activating factor), that is required for caspase activation. The survival-promoting function of ced-9 originally was represented by a single vertebrate homologue, bcl-2 (B-cell lymphoma-related gene), but subsequently other bcl-2 family members have been identified with similar anti-apoptotic functions (e.g., bcl-x), as well as other bcl-2 family members that are pro-apoptotic (e.g., bax, bak, bim). Loss-of-function mutations in mice by targeted gene deletion (gene knockout) of many of these vertebrate homologues have confirmed their role as important regulators of PCD.

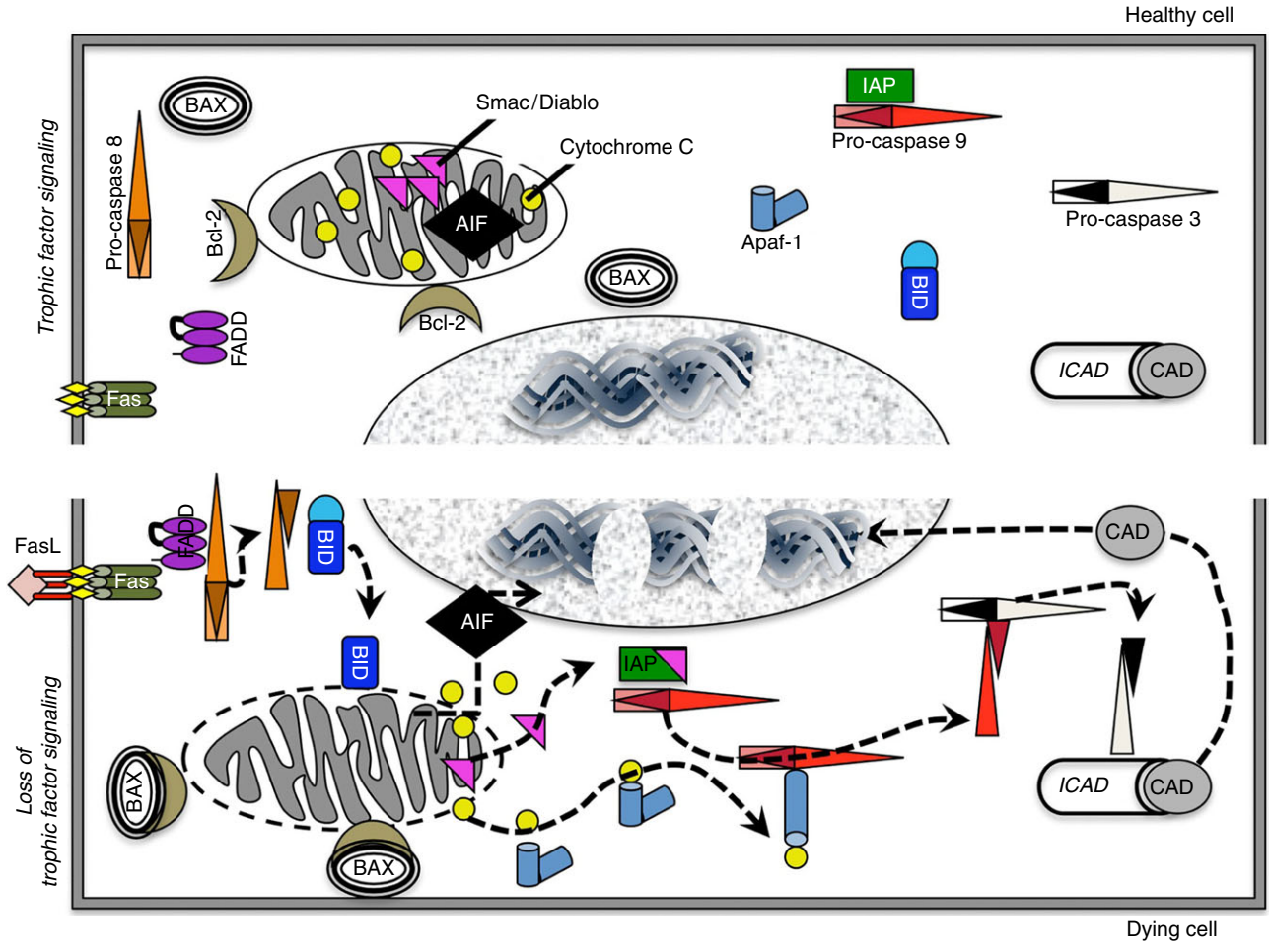
To understand cell death pathways, it is useful to consider the process of cell death in terms of signals that (i) induce death, (ii) regulate the intracellular process of cell death, and (iii) are involved in the removal of the corpse. NTFs act via membrane receptors resulting in intracellular signaling that either inhibits the expression or activity of pro-apoptotic genes or induces the expression or activity of anti-apoptotic gene products. On the other hand, as discussed below, more recent evidence indicates that pro-neurotrophins are capable of inducing cell death, suggesting that the availability of pro-survival versus pro-death signals can dictate the neuron's fate. Additionally, Fas-Fas ligand interaction has also been shown to regulate neuronal PCD.

Availability of NTFs induces or maintains activation of the PLCg, PKA, PI3-K/AKT, PKC, and ERK1/2 signal transduction pathways (see below). These pathways utilize both transcriptional and nontranscriptional mechanisms to regulate neuronal survival. For example, PI-3K/Akt has been reported to phosphorylate the pro-apoptotic molecule, Bad, promoting its association with 14-3-3 and preventing inactivation of pro-survival molecules, Bcl-2 and Bcl-x. Activation of these pathways also influences gene expression by activating transcription factors—namely cAMP responsive element binding protein (CREB)—which can result in transcription of anti-apoptotic genes like Bcl-2 and inhibitor of apoptosis proteins (IAPS). On the other hand, the JNK signal transduction pathway is activated in some neurons following NTF deprivation or Fas-Fas ligand interaction. JNK activation results in the activation of BH-3 proteins, regulation of Smac release from the mitochondria, and caspase activation.

Following inactivation of trophic receptors, and/or activation of p75NTR or Fas-Fas ligand, a series of intracellular events occurs that leads to a rapid and efficient



(A)



(B)

dismantling of the cell. A central integrator of cell death in vertebrates is the mitochondrion (Fig. 18.7). Both pro-survival and pro-apoptotic Bcl-2 proteins are critical for regulating mitochondrial membrane permeability. Permeabilization of the mitochondrial membrane results in increased reactive oxygen species, reduced ATP production, and release of the pro-apoptotic factors cytochrome c, Smac, and apoptosis inducing factor (AIF). Cytochrome c in the cytoplasm binds to Apaf-1, revealing Apaf-1's caspase recruitment domain (CARD), and in the presence of ATP forms a heptamer. Pro-caspase 9 has a high affinity for the CARD and binds at the

FIGURE 18.7 (A) Schematic representation of the major steps in the developmental PCD pathway of neurons in the nematode worm *C. elegans*. Both fly and mammalian homologues have been identified supporting the hypothesis that mechanisms of PCD are evolutionarily conserved. The mammalian homologues are listed in italics below the name of the *C. elegans* gene. (B) Some of the major events involved in mammalian PCD are illustrated. Healthy cells (top) receive signals to survive or lack signals to die (on left). Cell death mediators are present in healthy cells, but their location and/or association with regulators prevent activation of the cell death pathways. In cells that receive appropriate signals to die (bottom), cell death–specific pathways are activated. These pathways involve a permeabilization of the mitochondria membrane and release of factors that either directly or indirectly activate cell death–specific events. In neurons that fail to obtain NTF support (bottom), the pro-apoptotic gene Bax interacts with and inhibits the anti-apoptotic gene Bcl–2 in mitochondria. This results in the release from mitochondria of cytochrome-c, which forms a complex with Apaf-1 and caspase-9 that in turn activates downstream caspases such as caspase-3 that ultimately directly or indirectly degrade diverse nuclear and cytoplasmic targets. These degradative changes are what define apoptosis and result in eventual engulfment and phagocytosis of the apoptotic cell. In some situations two additional molecules released from mitochondria (along with cytochrome-c) are the pro-apoptotic proteins AIF, which can degrade the nucleus independent of caspases, and Smac/Diablo, which can inhibit IAP and promote the apoptotic pathway via caspase-9 and caspase-3. Although not shown here, in some situations developing neurons undergoing PCD activate cell cycle proteins that also serve a signaling function required for apoptosis (Copani et al., 2001). It is critical to note that all neurons do not use the identical pathways for cell death. For example, cell cycle proteins can mediate the death of sensory neurons; however, they do not play a role in motoneuron PCD (Taylor *et al.*, 2003. *Mol Cell Neurosci*. 24(2):323–39).

heptamer. This complex is referred to as the apoptosome. The increased local concentration of pro-caspase 9 results in its auto-cleavage, producing active caspase 9. Caspase 9 then cleaves pro-caspase 3 to the active form of caspase 3 that subsequently activates, inactivates, or destroys specific nuclear and cytoplasmic substrates, resulting in rapid cell degeneration. The activation of caspases as well as the availability of substrates may dictate the morphological changes associated with neuronal death (see previous section).

Many developing neurons appear to share the core cell death program described above, but there is also increasing evidence that alternative, nonapoptotic pathways may exist for mediating the normal PCD of some neurons. For example, when components of the apoptotic machinery are genetically deleted, cell death occurs by a morphologically distinct nonapoptotic death (Kroemer et al., 2009). Type II cell death or autophagy is characterized by the accumulation of double membrane–bound vesicles that fuse with lysosomes. Interestingly, Bcl-2 proteins and caspase activation appear to have regulatory roles in autophagy, suggesting that the different types of cell death utilize common mechanisms. Autophagy is also a critical nondeath event in maintaining homeostasis in many tissues, including the nervous system, and its role in normal cell death is still being debated. Another example of nonapoptotic cell death is necroptosis, a type of programmed necrosis recently characterized (Kroemer et al., 2009). Many of the stimuli that induce apoptosis instead activate necroptosis when caspases are not available. Morphological features of this type of death include organelle swelling, rapid mitochondrial dysfunction, plasma membrane permeabilization, and lack of nuclear fragmentation.

While Bcl-2 proteins and caspases are critical mediators of cell death, these factors also have important non–cell death–related roles. For example, Bcl-2 proteins are also involved in regulating mitochondrial fission and fusion independent of cell death. Caspases also have non–cell death roles, including proliferation, differentiation, and synaptic function (Fuchs & Steller, 2011). For example, caspase 3 is required for long-term depression and AMPA receptor internalization in hippocampal neurons.

The final stage of cell death is the removal of the corpse. This event is as critical as the other steps involved in the cell death process because ineffective removal of dying cells can result in apoptotic cells becoming necrotic and initiating inflammation or an autoimmune response. While nonprofessional phagocytes (e.g., ependymal or Schwann cells) can accomplish this task, professional phagocytes such as macrophages and microglia are most efficient at recognizing and removing dying cells (Mallat, Marín-Teva, & Chéret, 2005).

During early CNS development, microglia are usually not present in large numbers in the area of PCD. Some signal must be sent by the dying cells to recruit the phagocytes into the area. The phospholipid lysophosphatidylcholine, ATP, and UTP are macrophage chemotactic factors shown to be released by dying cells (Lauber et al., 2003; Elliott et al., 2009). Once the phagocytes are recruited to the area, they must be able to distinguish between healthy and dying cells. A series of evolutionarily conserved genes are involved in the expression of recognition markers on dying cells, expression of the receptors on phagocytic cells, as well as for those factors necessary for the processing of the engulfed apoptotic cells (Kinchen & Ravichandran, 2007).

Regulation of cell death is a critical issue in neurons because with only rare exceptions, postmitotic neurons are irreplaceable. One of these regulatory mechanisms is referred to as "competence to die." As neurons differentiate, they develop a dependence on NTF support. At this stage, in order to undergo PCD, developing neurons require not only the release of cytochrome c from mitochondria and caspase activation, but in addition, they require the loss of NTF support. In the presence of trophic support, the x-lined inhibitor of

apoptosis protein (XIAP) serves as a brake on caspase activation when cytochrome c is accidentally released; loss of NTF support removes this brake, thereby creating competence-to-die (Potts, Singh, Knezek, Thompson, & Deshinukh, 2003). Adult neurons also need to be protected from accidental cell death. As postmitotic neurons further mature past the period of naturally occurring cell death, they lose or reduce their dependence on NTFs such that they can now survive following NTF deprivation. This involves a block in the translocation of the pro-apoptotic Bax protein from the cytoplasm to the mitochondria, a failure of cytochrome c release and inhibition of apoptosis (Putcha, Deshmukh, & Johnson, 2000).

Summary

PCD is a metabolically active process that involves specific genetic pathways necessary for the cascade of events leading to degeneration. Several genes in the PCD pathway were first identified in *C. elegans* and homologues have been found in insects and vertebrates. The mitochondrion is a central integrator of cell death signaling. Although most developing neurons appear to share a common core PCD pathway, morphological and biochemical evidence also indicates the presence of alternatives to the core PCD machinery. One consequence of the increased evolutionary complexity in the regulation of neuronal PCD is the availability of additional checks and balances before a cell passes the irreversible point of commitment to die. Because virtually all neurons are postmitotic and neuron numbers finite, multiple, relatively fail-safe, mechanisms are required so that accidental death occurs only rarely.

NERVE GROWTH FACTOR: THE PROTOTYPE TARGET-DERIVED NEURONAL SURVIVAL FACTOR

Following the discovery of NGF by Levi-Montalcini, Cohen, and Hamburger in the 1950s and 1960s, and its possible role in the survival of developing sympathetic neurons, the characterization of the functional role of NGF became the archetypical model for the investigation of other putative NTFs. Analysis of the function(s) of a putative NTF involved several steps. Typically, the biological activities and target cell specificity of putative factors were first investigated *in vitro*. Primary cultures of neurons dissociated from peripheral ganglia were ideal for these early assays. Analysis of the developmental expression of specific NTFs and their corresponding receptors has been used to determine if both the ligand and the receptor are normally present at the appropriate time and place *in vivo* for the putative factor to serve in the regulation of a specific subpopulation of neurons.

In gain-of-function approaches, treatment of embryos with excess exogenous NTF has been used to determine if the survival or differentiation of responsive neurons is restricted by either the production or the access to a limited quantity of endogenous factor. Finally, in loss-of-function approaches, methods that inhibit or prevent the function of specific NTFs or their receptors have allowed investigators to determine whether the perturbation of endogenous NTF signaling alters normal development. These NTF/NTF receptor deprivation experiments have included treatment with activity-blocking antibodies that prevent trophic signaling, treatment with soluble receptor-derived antagonists that compete for and adsorb endogenous ligands, and the generation of transgenic mice with null mutations of either the factors or their receptors.

Sympathetic and sensory ganglia removed from developing animals have been shown to produce a dense halo of neurite outgrowth when treated with NGF. In fact, this "axonal halo" assay—not a neuronal survival assay—was the original biological activity first used to purify and characterize NGF as an NTF. NGF was shown to be required for the survival of dissociated sympathetic and some sensory neurons when they were grown in the absence of nonneuronal cells. This demonstrated that NGF could prevent cell death by the direct activation of receptors on isolated neurons. When developing embryos were treated with excess exogenous NGF, sympathetic and sensory ganglia were enlarged significantly, and neurite growth from these neurons increased markedly. In addition, ganglia in NGF-treated embryos contain many more neurons than normal because PCD had been prevented. The soma was enlarged significantly by NGF treatment, and the dendritic arbors of sympathetic neurons were more complex. These studies indicated that the supply or access to endogenous NGF in sympathetic and sensory targets was likely to be rate limiting for the survival and growth of these dependent populations.

The most convincing evidence that NGF is required for neuron survival has been gained from NGF deprivation experiments. Embryos treated with antibodies that selectively block NGF activity as well as the null mutation of either NGF or its trkA receptor in transgenic mice have both confirmed that sympathetic as well as some sensory neurons require NGF for survival (Fig. 18.8). NGF was localized to the peripheral targets of these neurons at the time of their normal innervation, consistent with its role as a target-derived survival factor. Further, the level of NGF synthesis was correlated with the density of target innervation, and the NGF receptor, trkA,

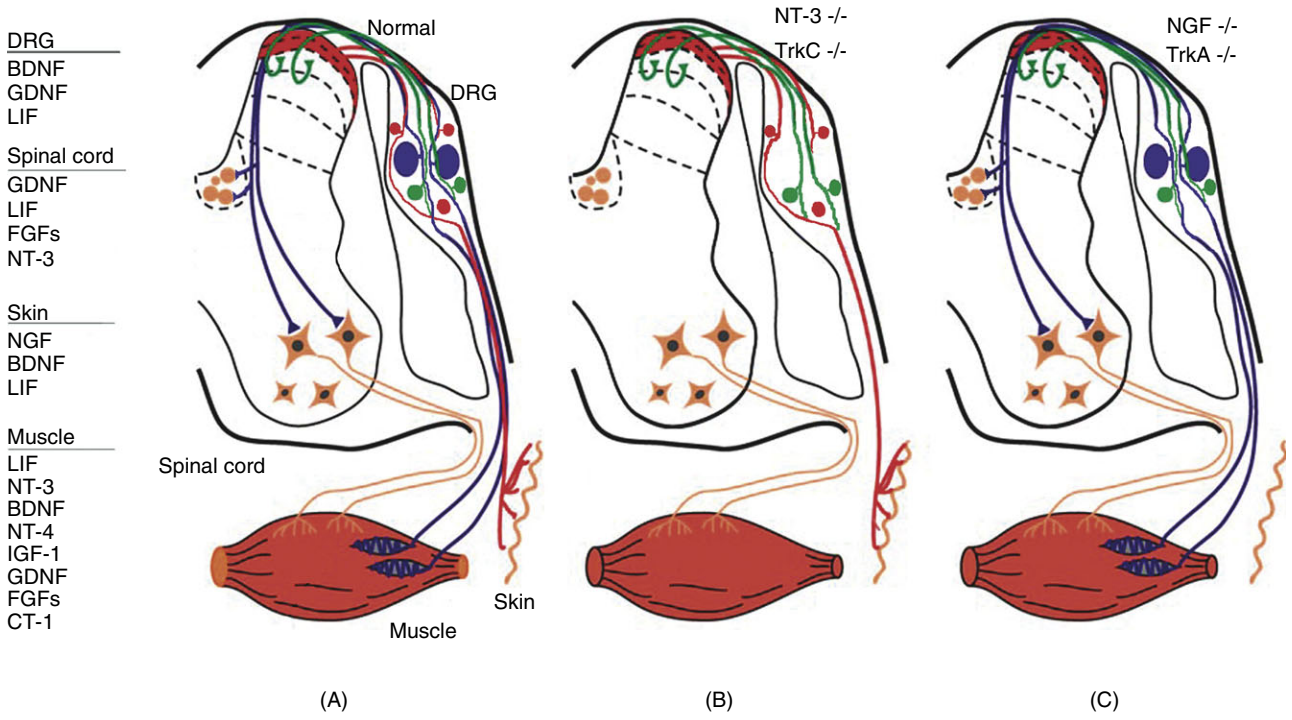


FIGURE 18.8 Phenotypic alterations in sensory/motor pathways caused by null mutations in NGF/trkA and NT-3/trkC. In the dorsal root ganglia (DRG) of normal mice (A), small-diameter (red), medium-diameter (green), and large-diameter (blue) neurons are present. Large-diameter neurons innervate muscle spindles and other proprioceptive end organs and have axon terminations in the lowest laminae of the dorsal horn and in the ventral horn. These neurons are lost when NT-3 or trkC is absent (compare A and B). Many of the small-diameter neurons innervate skin, respond to temperature and pain, and have terminations in the dorsal-most laminae of the dorsal horn. These neurons are lost when NGF or trkA is absent (compare A and C). DRG neurons indicated in green are neurotrophin-independent mechanoreceptive neurons (peripheral projections not shown). *Modified and redrawn with permission from Snider (1994).*

was localized to dependent afferent neurons at the times and places appropriate for regulating normal PCD. The localization of NGF synthesis in sympathetic targets and the loss of these afferent neurons with NGF deprivation firmly established NGF as a prototype target-derived NTF required for the survival of sympathetic neurons and a subset of sensory neurons.

One prediction of the neurotrophic hypothesis is that access to NGF in only the distant target region is adequate to support the survival of the remote cell body. This idea has been tested by Robert Campenot and colleagues *in vitro*. NGF has been applied selectively only to local axon terminals in a three-compartment tissue culture chamber to determine the long- and short-range effects of NGF treatment. NGF-dependent neuronal cell bodies in the central chamber survived when only their terminals were treated with the factor, indicating that target-derived NGF available only to neurites can generate and retrogradely transport the signal required for cell body survival. Peripheral processes were lost rapidly and selectively in outer chambers where NGF was withdrawn but were maintained and grew in the outer chambers where NGF was added. This important demonstration illustrates the capacity of target-derived NGF to have both direct long distance effects on the survival of neurons and direct local effects on the growth, maintenance, and sprouting of axonal branches. In addition, recent studies show that retrogradely transported NGF controls the assembly of afferent synapses on cell bodies and dendrites. NGF-responsive neurons possess both a "ligand-specific" receptor, trkA, and a "common" receptor, p75NTR (Fig. 18.9) that binds all neurotrophins with a similar affinity. Many of the biological activities of NGF have been attributed to the ligand-induced transduction of trkA.

Summary

The modern study of neuronal cell death began with investigations of how synaptic targets of sensory and motor neurons regulate their development. Viktor Hamburger and Rita Levi-Montalcini, beginning in the 1930s, ultimately showed that targets promote the survival and maintenance of innervating neurons. This notion, in turn, provided a conceptual framework for the discovery of the first target-derived NTF, nerve growth factor. From these beginnings, the neurotrophic hypothesis was formulated: neurons compete for limiting amounts of

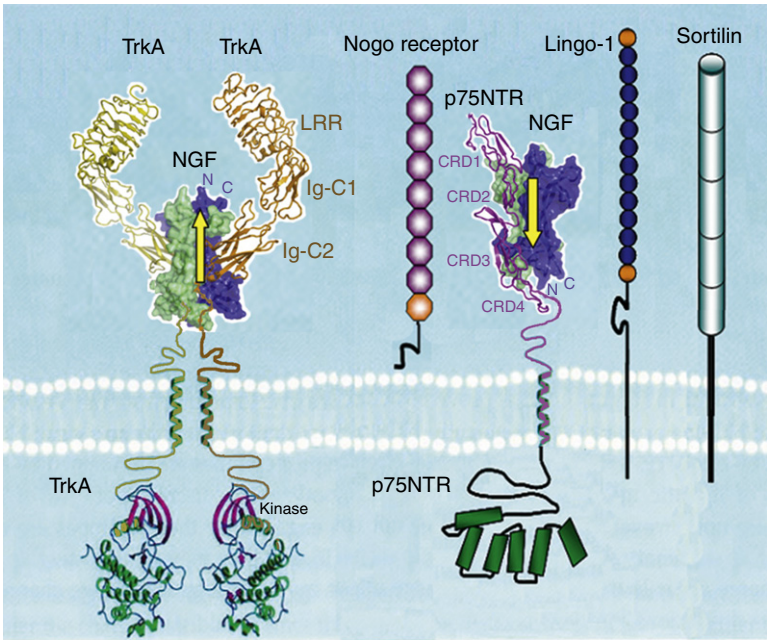


FIGURE 18.9 Models of NGF, the catalytic (full-length) TrkA receptor, and p75NTR and its binding partners (sortilin, Nogo receptor, and Lingo 1). Note the ability of one ligand (NGF dimer) to bring together two trk receptor molecules to initiate signaling. p75NTR lacks a cytosolic kinase domain, but associates with multiple partners (possibly not including direct interactions with trk receptors) to assemble signaling platforms. Note that the NGF dimer binds to a p75NTR monomer in an opposite orientation compared to TrkA (yellow arrows). C and N, C and N termini of NGF; Ig-C1, Ig-C2, C-terminal immunoglobulin ligand binding domains 1, 2; LRR, leucine-rich repeat; CRD 1–4, cysteine-rich domains 1–4. *Modified from Barker (2004) and Wehrman et al., 2007 (Neuron 53, 25-38) with permission.*

target-derived survival promoting (trophic) agents during development. NGF was established as the prototypical target-derived NTF.

THE NEUROTROPHIN FAMILY

Only a few subpopulations of peripheral neurons, including sympathetic and some sensory neurons, are exclusively dependent on NGF for survival during development. The survival of CNS neurons is largely unchanged following the null mutation of NGF. Other survival factors are now known to regulate neuron survival elsewhere in the nervous system. Extracts made from a number of tissues, as well as media containing proteins secreted by a variety of cultured neuronal and nonneuronal cells, have been shown to support the survival of many different classes of neurons that are not NGF dependent and that do not express the NGF receptor trkA, suggesting that additional NTFs exist. Because NTFs are made in extremely low quantities, the biochemical isolation of NGF-related molecules using conventional protein purification methods proved to be difficult.

A significant breakthrough occurred with the purification of a second NGF-related NTF by Yves Barde and colleagues in 1982. Unlike NGF, which was purified several hundredfold from an extraordinarily rich biological source (salivary gland) unrelated to the nervous system, endogenous brain-derived neurotrophic factor (BDNF) was purified several millionfold from adult pig brains. Each kilogram of starting material yielded only a microgram of factor that over time was eventually sequenced and cloned to produce recombinant factor. The molecular cloning and expression of BDNF opened the door for an accelerated period of research on NGF-related factors. When the protein structure of BDNF was compared with NGF, they were both found to encode homodimers of small, very basic secreted peptide ligands with an amino acid homology of approximately 50%. Using polymerase chain reaction primers prepared from homologous domains to search for other related proteins, investigators rapidly identified additional neurotrophin family members in multiple species (Lewin & Barde, 1996). Described as neurotrophins or nerve feeding factors (i.e., NT-3 and NT-4/5, NT-6), these additional proteins were cloned and sequenced without the requirement of exhaustive protein purification.

Each neurotrophin family member is synthesized as an approximately 250 amino acid precursor (proneurotrophin) that is processed into a roughly 120 amino acid protomer. The precursor was initially thought to be functionally irrelevant, but it turned out to have important death-signaling capabilities, as described below. Homologous regions of the several different family members are concentrated in six hydrophobic domains containing cysteine residues. The linkage formed by each homodimer ligand utilizes these regions to form a "cysteine knot" that maintains the twin protomers in juxtaposition. The secreted dimer appears as a symmetrical twin with variable regions containing basic amino acid residues exposed on the surface (Fig. 18.9). Because all family members share this core structure, they are remarkably similar, with three-dimensional symmetry around two axes. The symmetry of this twin structure allows the neurotrophin ligand to activate trk receptors by binding separate

receptor molecules together in the membrane, permitting docking of additional intracellular signaling molecules, and initiation of signal transduction cascades (Fig. 18.9). The exposed outer regions that vary among neurotrophin family members are responsible for receptor-binding specificity.

Summary

The purification, molecular cloning, and expression of BDNF opened a floodgate of research on an NGF-related family of NTFs called neurotrophins or nerve feeding factors. Each neurotrophin family member is synthesized as a pro-neurotrophin that is processed into a homodimer with a conserved region containing a cysteine knot in the core of the molecule. The secreted factor is a symmetrical twin with duplicate sites used for bivalent receptor binding and formation of signaling platforms.

NEUROTROPHIN RECEPTORS

NGF binds to a relatively small number of very high-affinity binding sites and a second set of about 10-fold more abundant, but lower affinity, binding sites at higher concentrations (Roux & Barker, 2002). A 75-kDa protein (p75) was purified and cloned first. This transmembrane glycoprotein with extracellular cysteine repeat motifs shares structural homology with the tumor necrosis factor receptor family. The cytoplasmic domain of p75 lacks the kinase domain present in most growth factor receptors for intracellular signal transduction, but it can associate with several other signaling proteins (discussed later) (Fig. 18.9). When expressed in fibroblasts, this receptor has low-affinity NGF-binding properties (ligand binding is rapidly on and off) and therefore also has been called the low-affinity NGF receptor. This name has proven to be a misnomer, since other neurotrophin family members also bind p75 with a similar affinity. It is therefore more appropriately named the "common" neurotrophin receptor (p75NTR).

A major breakthrough in the characterization of the NGF receptors came with the fortuitous discovery and cloning of an oncogene identified in a human colon cancer. The sequence of this 140-kDa transmembrane protein contained a cytoplasmic kinase common to many growth factor receptors. The corresponding proto-oncogene was named trk (pronounced "track"—for tropomyosin-related kinase). It rapidly was appreciated as a member of the tyrosine kinase-containing receptor superfamily with trk mRNA expression localized to NGF-responsive neurons.

Low-stringency screening of cDNA libraries with the original trk proto-oncogene probes led to the discovery of other related neurotrophin receptors. The NGF binding receptor was called trkA, whereas two additional 145-kDa members of this protein family were named trkB and trkC (Fig. 18.10). Expression of trkA conferred specific high-affinity NGF binding and NGF-induced receptor phosphorylation. NGF-signaling properties have been examined extensively in the NGF-responsive pheochromocytoma (PC12) cell line, derived from adrenal medullary cells. Mutant PC12 cell lines that have lost their capacity to respond to NGF contain many p75NTR receptors but lack trkA. Transfection of these mutant cells with trkA restores their biological responses to NGF treatment. The most convincing evidence for the necessity of trkA comes from the analysis of transgenic mice lacking functional trkA receptors. As expected, these mice have a phenotype that is similar to that of transgenic animals that have a null mutation for NGF (Table 18.2).

The trkB receptor is activated specifically by low concentrations of BDNF or NT-4/5 and, to a lesser extent, by higher concentrations of NT-3. NT-3 activates the trkC receptor most effectively. All trk receptors contain three leucine-rich motifs, two cysteine clusters, and two immunoglobulin-like motifs in the extracellular region, a transmembrane domain, and a tyrosine kinase domain in the cytosolic region (Fig. 18.9). The unusual combination of extracellular motifs makes up the ligand-binding region and places this family in a novel class of tyrosine kinase receptors. The region of highest sequence homology among family members and other growth factor receptors is in the kinase domain.

Receptor isoforms resulting from splice variants of trk mRNA transcripts exist for each family member. A variety of both full-length and kinase-deleted or truncated receptors are widely expressed on neurons throughout

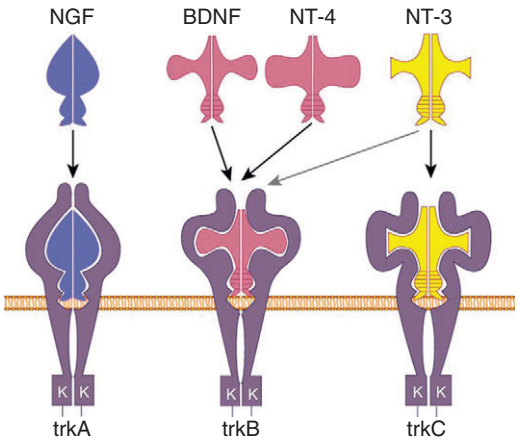


FIGURE 18.10 Ligand binding preferences of neurotrophins for each member of the trk receptor family. Not shown are the truncated (kinase deleted) isoforms of trkB and trkC. Other isoforms containing inserts and deletions also exist, providing a wide variety of receptors. NT-4 is also named NT-4/5. K = tyrosine kinase.

TABLE 18.2 Percentage of Neurons Lost in Neurotrophic Factor or Receptor Deficient Mice[a]

Neurotrophic Factor/Receptor Null Mutation	Viability	Losses in PNS Ganglia						Losses in CNS Nuclei	
		Dorsal Root Ganglia	Trigeminal Ganglia	Nodose Petrosal Ganglia	Vestibular Ganglia	Cochlear Ganglia	Superior Cervical Ganglia	Spinal Motoneurons	Facial Motoneurons
TrkA	Poor	–70–90%	–70–82%	?	–0%?	–0%	–95%	–0%?	?
NGF	Poor	–70%	–75%	–0–15%	?	–0%	–82–95%	–0%?	?
TrkB	Very poor	–0–41%	–39–60%	–90–94%	–56–85%	–15–20%	–0%	–0%[b]	–0[b]
BDNF	Moderate	–0–44%	–0–45%	–39–66%	–82–87%	–7%	–0%	–0%	–0–3%
NT-4	Good	–0–14%	–0–5%	–40–59%	–0–21%	?	–0%	–0%	–0–8%
BDNF/NT-4	Good	–0%	–9–34%	–79–90%	–82–90%	?	–0%	–0%	–0–11%
TrkC	Moderate	–17–38%	–21–48%	–14–18%	–15–29%	–50–85%	–0%	–0%	–0%
NT-3	Very poor	–36–79%	–61–68%	–30–47%	–15–34%	–85%	–48–53%	–35%[c]	–0%
TrkB/TrkC	Very poor	–41%	?	–95%	–58–100%	–61–100%	–0–4%	–0–5%	–0–5%
BDNF/NT-3	Very poor	–84%	–74%	–62%	–99%	–100%	?	?	?
BDNF/NT-4/NT-3	Very poor	–92%	–88%	–96%	–100%	?	–47%	–20%	–22%
gp130 (IL receptor	Very poor	–21%						–41%	–31%
CNTFRα	Very poor	–0%	–0%	?	?	?	–0%	?	–40%
LIFR	Very poor	?	?	?	?	?	?	–40%	–35%
c-ret	Very poor	–0%	?	?	?	?	–100%	–50%	–30%
GFRα1	Very poor	–0%	–0%	–15%	–0%	–0%	–0%	–24%	–0%
GFRα2	Poor	–0%	–0%	–0%	–0%	–0%	–0%	–24%	?
GFRα3	Good	–0%	–0%	?	?	?	–50–95%	?	?
GDNF	Very poor	–0–23%	?	–0–40%	–0%	?	–35%	–22–37%	–18%
Neurturin	Good	–0%	?	–0%	?	?	–0%	–0%	?
IGF-1	Variable							–30%	–25%

[a]*For references to original studies, see Huang and Reichardt (2001), Airaksinen and Saarma (2002), von Bartheld and Fritzsch (2006).*
[b]*Losses reported by Klein et al., 1993* (Cell **75**, *113–122) could not be replicated.*
[c]*Losses of gamma motoneurons only; Kucera et al., 1995 (Neuroscience 69, 321–330).*

the nervous system. Truncated receptors, which also are expressed on glial cells, can bind and internalize their cognate ligand and are capable of limited signaling, but they cannot initiate the phosphorylation events required for mainstream receptor signal transduction (see below). As a result, the distribution and membrane concentration of truncated receptors could potentially modulate neurotrophin activity.

Although trk receptors account for many of the biological responses of neurons to neurotrophins, p75NTR can modify trk ligand binding and neurotrophin specificity, it associates with additional signaling partners, and can initiate pathways for intracellular signaling independent of trk receptors (Fig. 18.9). Several mechanisms have been proposed to account for an accessory role of p75NTR. The fast on and fast off kinetics of p75NTR could maintain and increase the local concentration of neurotrophins at the membrane surface and thereby increase the access of trk receptors to ligands. The p75NTR does not appear to form a transient heterodimer with trk receptors, and NGF's binding sites overlap in a way that prevents them from binding simultaneously to p75NTR and trkA in a 2:2:2 complex (Fig. 18.9). Despite lack of evidence for a direct p75NTR/trk interaction, p75NTR may still "hand off" the factor for trk binding (Roux & Barker, 2002).

In addition to enhancing NGF binding to and activation of trkA, p75NTR initiates NGF responses in cells that lack trkA. NGF binding to p75NTR can cause activation and nuclear translocation of a transcription factor, nuclear factor kappa B (NFkB), that promotes cell survival. However, p75NTR is related structurally to members of the tumor necrosis factor receptor (TNFR) family, many of which regulate the onset of cell death programs in the immune system and, for that reason, are termed death receptors. The cytoplasmic domain of p75NTR contains a "death domain" sequence similar to active sequences found in the TNFR family, and the p75NTR mechanisms of inducing PCD appear to be shared with other death receptors.

In neurons that express p75NTR but not trkA, NGF or proNGF induce cell death via p75NTR binding (Roux & Barker, 2002). Once bound, p75NTR may activate a PCD signaling pathway via the Jun kinase cascade. The Jun kinase cascade activates the transcription factor p53 with gene targets that include the pro-apoptotic gene BAX. It should be noted that NGF and other trophic factors can be utilized in different cells or in different epochs of time to induce many diverse biological activities. It is noteworthy that under specific circumstances (such as the absence of trk activity), a factor originally identified for the capacity to prevent PCD instead is utilized to execute programs promoting neuronal death.

Neurotrophin receptors were generally considered to be inactive ("neutral" in terms of neuronal survival) unless bound by their ligand. However, as noted above, recent work has shown that several receptors, including trkA and trkC, actively transduce a negative (apoptotic) signal in the absence of the ligand, and thus behave as "dependence receptors" (Mehlen, 2010; Nikoletopoulou et al., 2010).

Summary

Many neurons have both specific and common binding sites for neurotrophins. Biological responses primarily are associated with high-affinity binding and rapid phosphorylation signaling events. All neurotrophins bind p75NTR, the common neurotrophin receptor. p75NTR lacks a cytoplasmic kinase domain but can independently initiate signaling, can associate with additional signaling partners, and can enhance signaling through trkA. There are three tyrosine receptor kinase (trk) family members: trkA, trkB, and trkC. Each receptor binds one or more members of the neurotrophin family. Splice variants of trks result in isoforms that include truncated receptors with reduced signaling capabilities. p75NTR is related to the TNFR family of death receptors and contains a cytosolic "death domain." Activation of p75NTR, especially by pro-neurotrophins, may serve to kill neurons or other cell types through well-established signaling pathways used to promote PCD.

SECRETION AND AXONAL TRANSPORT OF NEUROTROPHINS AND PRO-NEUROTROPHINS

Processing and Secretion of Neurotrophins

Neurotrophins are expressed as larger precursors that are cleaved by furins and prohormone convertases. The extent and location (intra- or extracellular) of cleavage differs between cell types. The precursor forms were believed for a long time to be of little functional significance, and it was thought that the mature neurotrophins were the only important moiety that was secreted. Research by Barbara Hempstead and her colleagues in the early 2000s demonstrated that many cells secrete pro-neurotrophins and that the release of both pro-neurotrophins and mature neurotrophins contributes to the spectrum of physiological functions. Pro-neurotrophins bind with higher affinity to p75NTR than mature neurotrophins and appear to be the "preferred" ligand for p75NTR. Whether endogenous pro-neurotrophins bind with lower affinity to trk receptors is still controversial. To activate different signaling pathways, the neurotrophin-bound p75NTR can associate with several signaling partners, including

Nogo receptor (a GPI-linked protein), sortilin (a member of the VPS10 family), and LINGO-1 (involved in myelin-based growth inhibition; Fig. 18.9). This complex is involved in regulating neurite outgrowth in response to myelin-derived proteins such as Nogo, MAG, and OMgp. Formation of these different platforms may explain the multiple effects of p75NTR in different cell types and contexts. Sortilin also facilitates anterograde axonal transport of trk receptors to the axon terminus, which is required for subsequent retrograde neurotrophin signaling.

The proteolytic cleavage of pro-neurotrophins, and thus the local balance and concentrations of pro- and mature neurotrophins, provides a mechanism to regulate the direction of action of neurotrophins in a wide range of cellular processes (Lu, Pang, & Woo, 2005). Recent work has shown that pro-neurotrophins can be released from neurons, apparently leading to elimination of either competing p75NTR-expressing neurons or neurites, in what may be considered a molecular correlate of neuronal competition (Deppmann et al., 2008; Singh et al., 2008).

Neurotrophic Factors and Synaptic Plasticity

The classic neurotrophic hypothesis postulates that NTFs, including neurotrophins, are released by postsynaptic target cells and bind to presynaptic receptors for internalization, retrograde axonal transport, and activation of signal transduction cascades after arrival at the neuronal soma (Fig. 18.1). Yet, NTF receptors are present on both post- and presynaptic sites of the synaptic cleft. *In vitro* studies have shown that NTFs, and in particular BDNF, have rapid effects on synaptic transmission (McAllister, Katz, & Lo, 1999; Poo, 2001). Numerous studies have confirmed that NTFs have important functions in synaptic plasticity, meaning that these "synaptotrophins" regulate the functional strength of synaptic transmission including long-term potentiation (LTP), an experience-dependent, long-lasting increase in chemical strength of neurotransmission that is a crucial mechanism in learning and memory (Table 18.3).

The regulation of secretion of neurotrophins is particularly important for mechanisms of synaptic plasticity (Lessmann, Gottmann, & Malcangio, 2003). Neurotrophins can be released by constitutive modes (from cell bodies and processes) and by activity-dependent modes of secretion (from dendrites and axons). The pattern of activity, such as high-frequency stimulation, is crucial for the release. LTP is relatively synapse-specific; only active and successfully transmitting synapses are strengthened, so the regulated release of neurotrophins and subsequent signaling should be restricted to active synapses. Depolarization and cAMP-dependent pathways are thought to regulate the locally restricted release of neurotrophins such as BDNF. This may contribute to the local "tagging" or marking of the active synapses destined for strengthening. Trafficking of BDNF mRNA within the neuron might also contribute to synapse-specific plasticity, if the mRNA is "captured" for local translation at active synapses. Interestingly, a naturally occurring polymorphism of the BDNF gene (a Val-Met substitution in the 5'pro-region of BDNF) gives rise to a reduced capacity for activity-dependent (but not constitutive) secretion of BDNF, causing abnormal hippocampal function and some learning and memory deficits in humans (Egan et al., 2003).

TABLE 18.3 Nonsurvival Functions of Neurotrophic Factors[a]

Function	Examples of Relevant Trophic Factors and Cytokines
Cell fate decisions (stem cell/neuronal precursor derivatives)	BDNF, NGF, EGF, NT-3, CNTF, LIF, oncostatin M, CT-1
Axon guidance	NGF, BDNF, HGF, GDNF, neurturin
Neurite growth, branching	BDNF, NT-3, NT-4, CNTF, CT-1, HGF, GDNF, LIF, IGF-I
Myelination	IGF-1, BMPs, NT-3, BDNF, GDNF, CNTF
Dendrite development	BDNF, NT-3, NGF, GDNF
Synapse development	BDNF, NT-3, GDNF, neurturin, NT-4, CNTF
Synapse stabilization and plasticity	BDNF, NT-3, GDNF, FGFs, HGF, VEGF, CNTF, LIF, NGF, IGF-I
Long-term potentiation, learning, and memory	BDNF, NT-3, NGF, NT-4, IGF-I, interleukins
Pain physiology and pathophysiology	NGF, BDNF, GDNF, artemin, NT-3
Neurotransmitter phenotype	NGF, LIF, FGFs, GDNF, CNTF, BDNF, NT-3, IGF-I
Ion channel regulation/function	BDNF, interleukins
Neurological disease and neurodegeneration	
Seizure, epilepsy	BDNF, NGF, NT-3, VEGF, GDNF, LIF, CT-1, oncostatin M, CNTF
Stress, ischemia	NGF, NT-3, BDNF, IGF-I, VEGF, GDNF, HGF, BMPs
Traumatic injury, axonal regeneration, inflammation, stress	GDNF, IGF-I, LIF, CNTF, oncostatin M, CT-1, NGF
Psychiatric disease: depression	BDNF, NGF, FGFs, IGF-I, GDNF

[a]*For references, see Snider (1994), McAllister et al. (1999), Huang and Reichardt (2001), Poo (2001), Chao (2003), Zweifel et al. (2005).*

Endogenous neurotrophins may be derived from either pre- or postsynaptic sites, and they may act on either side of the synaptic cleft in paracrine or autocrine modes to implement functional and structural changes in synaptic plasticity. Trophic responses can also be modified by the recruitment of receptors to surface membranes from internal storage sites. Furthermore, after release and binding of newly synthesized neurotrophins, they can be recycled, and the internalized neurotrophin is not necessarily degraded immediately after the initial binding and internalization steps. Recycled neurotrophins have been shown to play functionally significant roles in LTP.

The Endosome Signaling Hypothesis of Retrograde Axonal Signaling

The neurotrophic hypothesis postulates retrograde axonal movement of trophic signals. Although evidence for retrograde transport of exogenous neurotrophins (NGF) was established in the 1970s, the precise mechanisms of axonal transport of trophic signals from the nerve terminal to the neuronal soma remained largely elusive. The prevailing hypothesis is that neurotrophins bind presynaptic surface membrane receptors on nerve terminals, are internalized, and form a ligand-receptor complex with the neurotrophin inside transport vesicles, leaving the trk receptor with its kinase domain on the outside of the vesicle or endosome (Fig. 18.11). The trk receptors may remain "active" (phosphorylated) during axonal transport along microtubules, because the ligand remains bound to the receptor within the vesicle. When the endosome, with the ligand-receptor complex attached, arrives at the neuronal soma, the phosphorylated trk receptor initiates a signal transduction cascade that results in changes of gene expression. Signaling endosomes can be isolated from axons; trk receptors and other signaling proteins interact at the level of the axon; trks bind to the dynein light chain subunit (part of the dynein motor complex) directly, and thus partake in the axonal transport machinery (Fig. 18.11). Furthermore, axon-derived endosomes that reach the soma may generate a different signal (activate a different ERK) than those generated by binding to cell surface receptors on the soma.

However, not all trophic signals from the nerve terminal that arrive at the soma may require the neurotrophin to be physically transported in a ligand-receptor complex. In addition to retrograde axonal transport from nerve terminals to the soma, trophic signals also are transported anterogradely along axons (possibly even a quantitatively more important route) from soma to nerve terminals, mediated by kinesin motors, followed by activity-dependent release of neurotrophins from axon terminals.

Summary

Expression and secretion of pro-neurotrophins has significant physiological functions in the regulation of cell death. Pro-neurotrophins bind preferentially to the p75NTR. The p75NTR associates with several different signaling partners to form multiple signaling platforms that include sortilin and Nogo receptors. Secretion of neurotrophins requires activity-dependent stimulation patterns that may specifically strengthen active synapses. Neurotrophins are transported retrogradely along axons in signaling endosomes, but they can also be transported anterogradely for release from axon terminals. Trk receptors are expressed both pre- and postsynaptically, forming paracrine as well as autocrine loops of synaptic signaling.

SIGNAL TRANSDUCTION THROUGH TRK RECEPTORS

Neurotrophin binding to trk receptors at the cell surface causes the formation of receptor dimers and coactivation of their tyrosine kinase activity. The homodimeric structure of the factors allows each bivalent ligand to bring two separate receptor molecules into close proximity. Aggregated receptors phosphorylate each other on specific tyrosine substrates within intracellular domains. The generation of phosphotyrosine residues in turn activates the receptor kinase and further catalyzes the formation of large signaling complexes through the recruitment of adaptor proteins. These proteins link the activated receptor kinase with intracellular signaling pathways shared by other growth factors (Ip & Yancopoulos, 1996). Once activated, the receptor initiates intracellular signals both locally in the cytoplasm and by a series of enzymatic cascades that eventually produce changes in gene transcription within the nucleus (Segal & Greenberg, 1996).

Receptor signal transduction involves multiple signaling pathways that can differ between individual neurons or even in the same neurons depending on recent events. Thus, the response of neurons to trophic factor stimuli is dependent on the current intracellular status of the cell in a dynamic fashion. The molecular components of these pathways are so well conserved that many of the signaling proteins are interchangeable among invertebrate and vertebrate species.

Three major pathways have been identified to implement trk signal transduction events that mediate survival as well as many other cellular responses to neurotrophins (Fig. 18.12). They are (1) the *phospholipase C (PLC-g) pathway*, (2) *Ras-MAP kinase pathway*, and (3) *phosphatidylinositol-3 kinase (PI3K) pathway*. The latter two pathways begin with adaptor proteins that contain a structural motif, the src

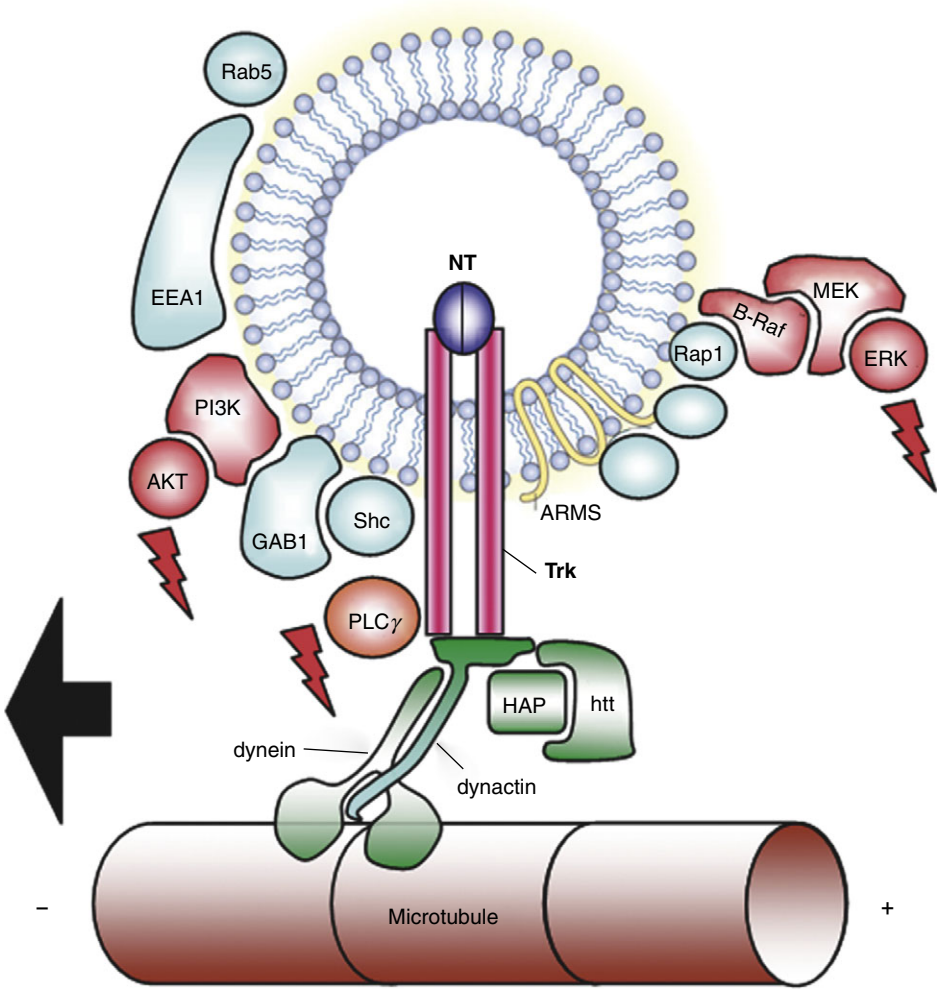


FIGURE 18.11 The signaling endosome. The internalized neurotrophin (NT)–receptor (Trk) complex is transported retrogradely from the nerve terminal to the cell body. The internalized vesicle serves as a platform for docking of adaptor proteins (light blue) that elicit signal transduction cascades via kinases (red); molecular motors and regulators (green) attach for transport along microtubules (brown). The arrow indicates the direction of transport toward the minus end of microtubules. The main signal transduction cascades are via the phospholipase C-γ (PLCγ), Raf-MAPK-ERK, and PI3K signaling pathways (red lightning symbol). For details of signal transduction, see legend to Figure 18.12. Rap1 and Rab5 are small G-proteins. Other abbreviations: Akt, v-akt murine thymoma viral oncogene homologue (protein kinase B); ARMS, ankyrin-rch membrane spanning protein; B-Raf, v-raf murine sarcoma viral oncogene homologue B1; EEA1, early endosome antigen 1; ERK, extracellular signal-regulated kinase; GAB1, GRB2 (growth factor receptor bound protein 2)-associated binding protein1; HAP, huntingtin-associated protein 1; htt, huntingtin; MEK, MAPK (mitogen-activated protein kinase); NT, neurotrophin; PI3K, phosphatidylinositol 3-kinase; Shc, Src homology 2 domain-containing transforming protein C; Trk, trk tyrosine kinase receptor. For details and references, see Zweifel, Kuruvilla, and Ginty (2005) and Gauthier et al. (2004).

homology domain 2 (SH2), which specifically recognizes the phosphotyrosine residue and flanking sequences. PLC-g activity generates two distinct second messenger signals: inositol triphosphate (IP_3) and diacylglycerol (DAG). IP_3 rapidly releases intracellular Ca^2 sequestered in local membrane compartments. This signal initiates the activity of local Ca^2 dependent enzymes (protein kinases and phosphatases). Similarly, DAG regulates the activity of DAG-dependent enzymes. Both Ras and PI-3 kinase pathways are engaged via adaptor and associated linker or extender proteins.

One of the best known Ras-dependent signaling pathways involves the ERK family of MAP kinases. This cascade is composed of serine/threonine kinases that are serially phosphorylated and activated. The initial member of this cascade is Raf, which phosphorylates and activates MEK, which phosphorylates the MAP kinases (ERK1 and ERK2). Once activated, these MAP kinases in turn phosphorylate a number of cytoplasmic and nuclear effectors. Importantly, activation of MAP kinases and their substrate, the protein kinase RSK, results in the phosphorylation of a number of transcription factors, including CREB (cAMP responsive element-binding protein), which controls expression of immediate-early genes (i.e., *c-fos* and *c-jun*) and delayed response genes. Once activated, transcription factors cause rapid and long-lasting changes in gene expression regulating cell survival, axonal and dendritic growth, neuronal differentiation, synaptic potentiation, and plasticity. In addition, the requirement on Ras-dependent pathways for biological responses varies among neuron types.

Neurotrophic factor receptor activation of the PI-3 kinase pathway is essential for the normal survival of many neurons. The phosphatidyl inositides made by PI-3 kinase regulate in part the activity of AKT (protein kinase B). This important protein kinase plays a critical role in controlling the biological activity of several regulatory proteins that govern normal PCD (see Fig. 18.12 and previous sections). Trk receptor activation of both the PI-3 kinase and the Ras-ERK pathways suppresses the capacity of activated p75NTR to induce cell death programs via Jun kinase. p75NTR may therefore induce neuronal death when appropriate neurotrophin binding to trk receptors is absent. Interactions between the multiple trk receptor signaling pathways in different cell

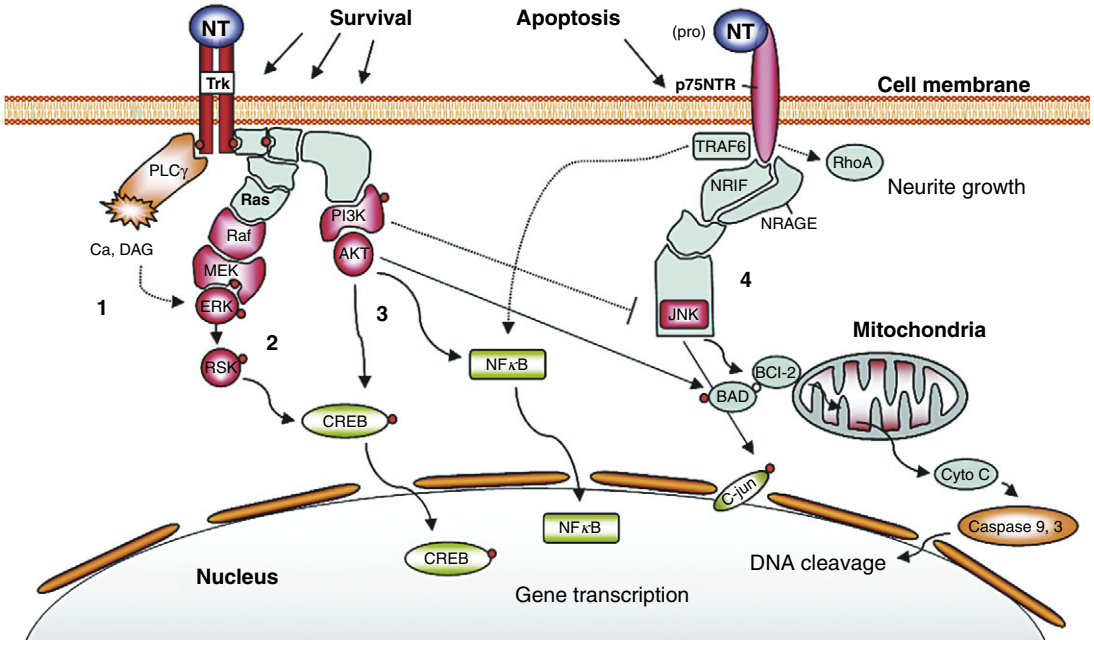


FIGURE 18.12 Trk and p75NTR signaling pathways. Neurotrophin (NT) dimers bind two trk receptor monomers or one p75NTR to activate survival (1–3) or apoptotic (4) signaling pathways. Ligand binding initiates trk receptor transduction by the phosphorylation (indicated by small red dot) in the cytoplasmic domains. The activated Trk kinase docks adaptor and linker proteins (light blue), which engage signaling cascades often containing multiple kinases (red). The three principal signaling pathways illustrated are the (1) phospholipase C pathway, (2) Ras-MAP kinase pathway; and (3) PI-3 kinase pathway. These pathways lead to nuclear translocation of transcription factors (green), such as CREB and NFκB, and ultimately regulation of gene expression. The binding of neurotrophins or proneurotrophins to p75NTR can activate BAD via the JNK cascade and eventually engages, via release of effectors from mitochondria, caspases involved in apoptosis (see also Fig. 18.7). There are several venues for crosstalk between the survival- and pro-apoptotic signaling cascades as indicated. Abbreviations not already explained in Legend 18.11: BAD, bcl2-antagonist of cell death; Bcl2, B-cell lymphoma 2; Ca, calcium; CREB, cyclic AMP response element binding protein; Cyto C, cytochrome C; DAG, diacylglycerol; JNK, C-Jun N-terminal kinases; NFκB, nuclear factor-kappa B; NRAGE, neurotrophin receptor interacting MAGE (melanoma antigen gene expression) homolog; NRIF, neurotrophin receptor interacting factor; p75NTR, p75 neurotrophin receptor; (pro)NT, pro-neurotrophin; Raf, Ras, and RhoA, small GTPases; RSK, ribosomal S6 kinase. For details, see text, and Huang and Reichardt (2001), Chao (2003), and Barker (2004).

phenotypes and under a variety of dynamic receptor signaling conditions provide a rich assortment of interrelated mechanisms to govern the PCD machinery.

Summary

Three tyrosine receptor kinases and trk family members—trkA, trkB, and trkC—transduce neurotrophin signals. Signaling pathways used by the trk family members are shared with those activated by many other growth factor receptors. Neurotrophin binding to trk causes receptor dimerization and phosphorylation of cytoplasmic tyrosine residues. Phosphotyrosines recruit cytosolic adaptor proteins that couple the activated receptor with intracellular signaling pathways. The biological response of the cell to a neurotrophic factor is dependent on the dynamic status of the pathway that can vary with recent cell history. Three of the well-known signaling pathways are PLC-g, Ras-ERK kinase, and PI-3K. Normal PCD is governed in many neurons by PI-3 kinase activation of the protein kinase Akt. Trophic factor deprivation of dependence receptors can activate pro-apoptotic mechanisms that promote cell death.

MULTIPLE CYTOKINES AND GROWTH FACTORS IMPLEMENT NEUROTROPHIC ACTIVITIES

Several Cytokines Mediate Cell Interactions within the Nervous System

We have emphasized neurotrophins as an example of the activities of NTFs in the developing and mature nervous system. Yet there are several other important families of NTFs. Often NTF families utilize receptors with co-receptor complexes (e.g., neurotrophins—trk, p75NTR; and GDNF family ligands—ret and GFRs). In many tissues, intercellular induction factors are important for governing the proliferation and differentiation of both embryonic and adult stem cells. In addition, these factors also play an important role in the response of tissues to trauma, inflammation,

infection, or tumor growth. In 1974, Stanley Cohen proposed that both lymphocyte-derived and nonlymphocyte-derived chemotactic and migration inhibitory factors be grouped into families of cytokines ("cell movement factors"). More recently, this term has been adopted as a general umbrella for many families of secreted proteins that mediate diverse biological responses, including changes in the immune system (interleukins), tumor cytotoxicity (tumor necrosis factors), and inhibition of viral replication or cell growth (interferons) (Table 18.4).

Many cytokines originally were named according to the particular biological activity that was utilized for their isolation, only to be later rediscovered or identified as important mediators of other physiological processes. For example, some factors were isolated on the basis of their ability to enhance the survival of specific populations of neurons isolated *in vitro*. These include ciliary neurotrophic factor (CNTF); glial cell line–derived neurotrophic factor (GDNF); and the other GDNF family members, neurturin, persephin, and artemin. CNTF is a member of a broader family of neuropoietic cytokines, including leukemia inhibitory factor (LIF), oncostatin M, cardiotrophin-like cytokine/cytokine-like factor (CLC/CLF) and cardiotrophin-1 (CT-1), that share a common three-dimensional structure and receptor subunits. CNTF originally was isolated and named because it supports the survival of neurons cultured from the parasympathetic ciliary ganglion.

The GDNF family primarily supports enteric neurons, dopaminergic neurons, and some motoneurons. Other proteins that originally were identified as mitogens or chemotactic factors in nonneuronal tissues also have been shown to affect either the survival or the differentiation of neurons, including the fibroblast growth factors (FGFs), insulin-like growth factors (IGFs), and hepatocyte growth factor (HGF). More recent additions to this list include cerebral dopamine neurotrophic factor (CDNF) and mesencephalic astrocyte-derived neurotrophic factor (MANF) (Lindholm & Saarma, 2010). The first neuronal function identified for LIF was the induction of cholinergic properties in cultured sympathetic neurons, but it also supports the survival of several classes of neurons and induces neural precursors to become astrocytes.

LIF, which has a number of actions in the immune system and other nonneuronal tissues, shares receptor subunits with the CNTF family and therefore can mimic both the survival and the cholinergic differentiation activities of CNTF observed in culture. Many cytokines important for the development or maintenance of other organs and tissues are also widely expressed within the nervous system. Their roles in the nervous system, however, remain to be defined. Likewise, factors first recognized as neuronal survival factors have mitogenic properties for either nonneuronal cells or neuronal precursors. As a result, these pleiotropic factors are grouped into families based on their protein sequences and receptor usage rather than on their biological properties. Table 18.4 summarizes major known ligands, receptors, and biological functions for cytokines and growth factor families.

Neurotrophic Factors Have Multiple Activities

Neurotrophic factors that prevent neuronal death during development appear to have many other important biological activities, including effects on cell

TABLE 18.4 Cytokine and Growth Factor Families

Family	Representative Members	Original Biological Activities
Neurotrophins	NGF, BDNF, NT-3, NT-4/5, NT-6	Neuronal survival and differentiation
Neuropoietic cytokines	CNTF, LIF, CT-1, CLC/CLF, oncostatin M	Survival of ciliary neurons and motoneurons, leukemia inhibitory activity, increased cholinergic properties
Tissue growth factors	TGF-α, TGF-β, FGFs, IGF-Iα,	Cell proliferation and differentiation in diverse tissues and organs
GDNF family	IGF-Ib, IGF-II, EGF, PDGF GDNF, neurturin, artemin, persephin	Neuronal survival and dopaminergic cell differentiation, and morphogens in kidney and sperm development
Interleukins	IL-1α, IL-1β, IL-2 through IL-15	Immunoregulation, diverse activities in the immune system
Tumor necrosis factors	TNF-α, TNF-β	Tumor cytotoxicity
Chemokines	MCAF, MGSA, RANTES, NAP-1, NAP-2, MIP-1	Leukocyte chemotaxis and cell activation
Colony-stimulating factors	G-CSF, M-CSF, GM-CSF	Hematopoietic cell proliferation and differentiation
Interferons	IFN-α, IFN-β, IFN-γ	Inhibition of viral replication, cell growth, or immunoregulation
CDNF/MANF family	CDNF, MANF	Survival of dopaminergic neurons

proliferation, migration, differentiation, axonal growth and sprouting, alterations in dendritic arbors, and synaptic plasticity of the nervous system (Table 18.3). Frequently, different populations of neurons respond to the same factor in distinct ways, and the same neuron may respond differently to the same factor at different developmental stages. Variations in neuronal responses to the same factor appear to depend not only on modifications of trophic receptors and their binding partners, but also on potential differences in the intracellular context of downstream signaling pathways. The distinctive activities of an NTF on different cells or during different epochs of development reflect the intrinsic properties of differentiating neurons or a dynamic change in a trophic response due to alterations in neural activity or other signaling events.

As described earlier, neurotrophins have been shown repeatedly to play fundamental roles in the survival of many peripheral neuronal populations (Table 18.2; Fig. 18.8). However, relatively few changes in the number of neurons have been observed in the CNS of neurotrophin or trk receptor null mutant mice. This observation was surprising because significant cell death occurs in the CNS, and NTFs and their receptors are widely expressed in the CNS during this period. The complexity and number of synaptic relationships established by CNS neurons compared to those established by PNS neurons may partially explain why CNS neurons are not as sensitive to the loss of a single neurotrophin or neurotrophin receptor (Fig. 18.13). The trophic support to CNS neurons is likely to arise from multiple families of NTFs with possible synergistic and/or compensatory effects.

The traditional view of the neurotrophic hypothesis has been that trophic support is derived from target tissues, but other sources of trophic support are now recognized (Figs. 18.13 and 18.14). Interpretation of mRNA localization studies is complicated by the fact that NTFs may not simply be released locally in the region of the cell body, but may also (or instead) be delivered to distant regions of the nervous system by anterograde or retrograde axonal transport. Thus, although the original neurotrophic hypothesis that neurons depend on target-derived trophic factors still holds true for many neurons during critical periods of development (including sympathetic neuron dependence for target-derived NGF), other sources of trophic support appear to play an important functional role in development, as briefly described

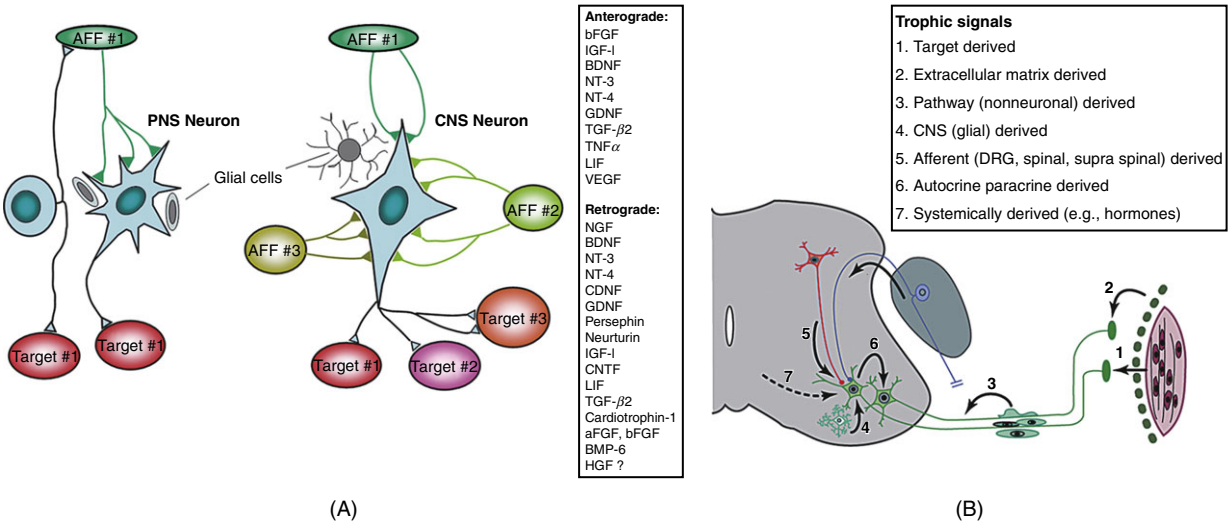


FIGURE 18.13 Possible sources of trophic support for peripheral (PNS) and central (CNS) neurons. (A) Many peripheral neurons such as sympathetic and bipolar sensory neurons (left side) have only two sources of support: one in the periphery (Target #1) for retrograde support and one from afferents (AFF #1) or the central target (for bipolar sensory neurons). Glial cells (gray) may also provide trophic factors. In contrast, central neurons (right side) receive synaptic input from many different types of neurons (AFF #1, 2, and 3), which may serve as a source of anterograde trophic support. Central neurons may also project to several different targets (Targets #1, 2, and 3), which each may provide retrograde trophic support. We list trophic factors that have been demonstrated to be either anterogradely or retrogradely transported along axons, with question marks when transport is suspected, but not proven. *Adapted, with permission, from Snider (1994) and von Bartheld, Wang, and Butowt (2001).* (B) Schematic illustration of different sources of potential trophic signals acting on one specific cell population (motor neurons in the spinal cord). Motor neurons (green cell bodies) can receive trophic support from a number of different sources. Axon terminals of the motor neurons have access to diffusible muscle-derived (1) or extracellular matrix-associated (2) trophic factors. Schwann cells (3) in the peripheral nerve or ventral root are another source of trophic support. Glial cells (4) in the spinal cord (astrocytes and/or oligodendrocytes) may also influence motor neuron survival. Motor neurons receive afferent input from several sources, including descending fibers from the brain, spinal interneurons, and axons from dorsal root ganglion (DRG) neurons (5), which can supply trophic support. Finally, motor neurons may influence their own growth and survival via autocrine or paracrine trophic support (6), as well as respond to trophic support provided by circulating hormones (7).

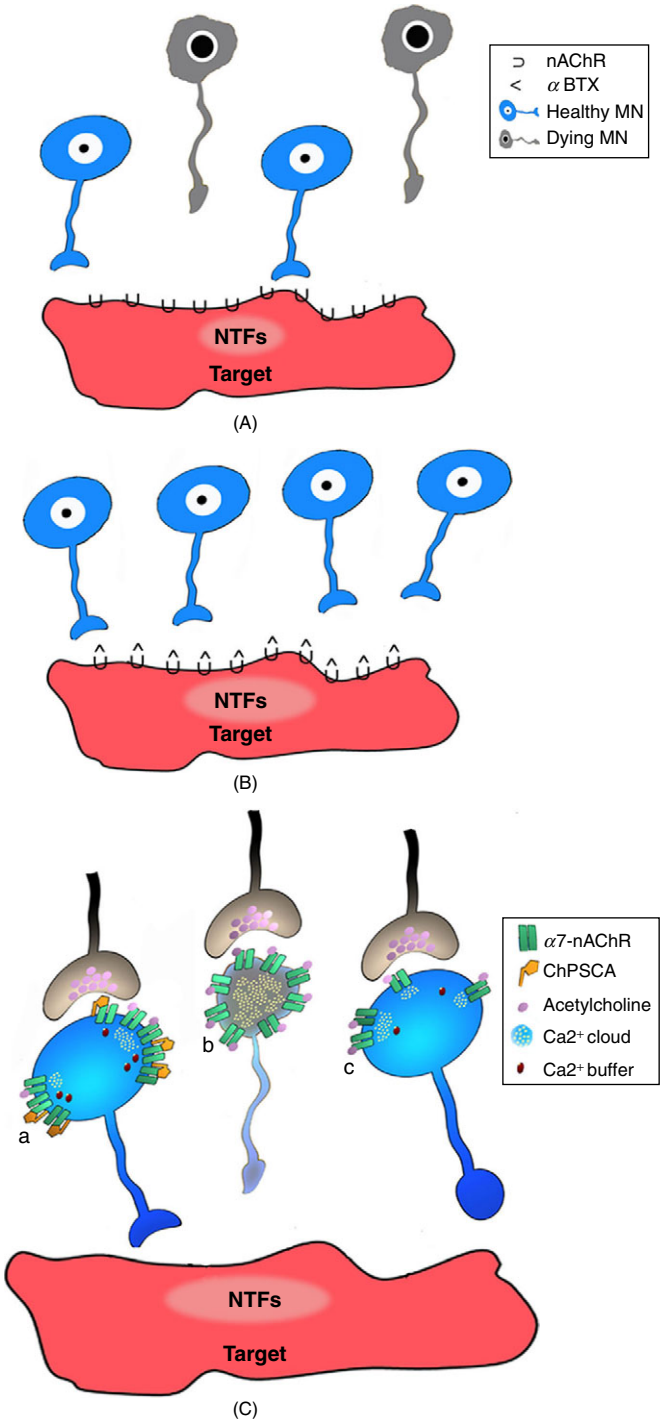


FIGURE 18.14 During embryonic/fetal development motoneurons establish functional synaptic connections with skeletal muscles via the neuromuscular junction (NMJ). This normal synaptic transmission activates skeletal muscles resulting in embryonic movements. If left unperturbed, this muscle activity reduces the production of (or access to) muscle-derived NTFs resulting in the PCD of a subset of motoneurons (MNs) (A). By contrast, if muscle activity is blocked, either genetically or pharmacologically, causing paralysis, the PCD of MNs is prevented (B). In addition to the role of synaptic activity at neuronal targets (e.g., the NMJ) in regulating PCD, afferent synapses also play a role in regulating PCD. A model is illustrated that integrates afferent- and target-derived influences on the choice of neuronal death or survival (C). In the developing chick embryo ciliary ganglion (CG, blue cells), young CG neurons are very heterogeneous with respect to the number of α7nAChRs that regulate calcium influx (e.g., b vs c) and they are already innervated by afferent cholinergic axons from the accessory oculomotor nucleus (gray terminals). There is also very little of the endogenous chicken prostate stem cell antigen (chPSCA), a prototoxin that blocks α7nACHR activation (b). Thus, young neurons with excessive α7nAChRs have a higher probability of PCD. As the developing neurons grow out axons that reach their targets in the eye, they become exposed to target-derived NTFs such as CNTF, GDNF, activin as well as other as yet unidentified factors. The influence of these factors facilitates maturation of the neurons, enhancing calcium buffering as well as inducing the expression of chPSCA. The end results are mature neurons (a), stable synapses, a population of silenced nAChRs together with active nAChRs at the stabilized synapses (*Part **C** courtesy of Rae Nishi; see Hruska and Nishi, 2007 (J. Neurosci.* **27**, *11501-11509)*).

earlier. Additionally, as described below, neuronal survival not only depends on competing successfully for trophic support but also on avoiding those cell–cell interactions that can actively kill cells, another example of how the traditional view of the neurotrophic hypothesis has had to be modified.

Finally, although beyond the scope of the present chapter, neurotrophins and other NTFs and their receptors also are distributed throughout the mature brain, and NTFs can alter neural activity by rapid and long-term changes in synaptic transmission. Neurotrophins and other NTFs also play a role in modulating long-term changes in functional and anatomical plasticity in the developing and mature brain by altering long-term potentiation, synaptic connectivity, development of adult-generated neurons, and responses to stress, inflammation, and trauma (Table 18.3).

Summary

There are multiple families of NTFs. CNTF and LIF belong to a neuropoietic cytokine family. The expression and biological properties of these cytokines distinguish them from neurotrophins. They possess widespread neurotrophic activity for many different neuronal and nonneuronal populations *in vitro* and facilitate the cholinergic differentiation of sympathetic and motor neurons. The GDNF family (GDNF, neurturin, artemin, and persephin) is related to the TGF-β gene family and has effects on enteric, dopaminergic, and motor neurons. FGFs may play important roles during development or after injury. NTFs have different functions

during different developmental stages: After early effects on proliferation and phenotype determination (stem cells), NTFs influence neuronal migration, survival, differentiation, and dendritic, as well as axonal growth. At later stages, NTFs regulate synaptic competition and plasticity, LTP, responses to injury, regeneration, and adult neurogenesis. The activity of multiple trophic factors from multiple sources is integrated to achieve a functional circuitry.

PROGRAMMED CELL DEATH IS REGULATED BY INTERACTIONS WITH TARGETS, AFFERENTS, AND NONNEURONAL CELLS

The PCD of vertebrate neurons and their precursors can occur at any stage of neuronal development from neurogenesis and neurulation to the time of establishment of synaptic connections with targets and afferents, and can involve mitotically active cells and migrating neurons, as well as undifferentiated and immature postmitotic cells (Fig. 18.3). However, the most common and historically the best studied type of PCD of neurons involves postmitotic, differentiating cells that die while establishing synaptic connections with other neurons and target cells (Oppenheim, 1991).

Because massive neuronal death and its regulation were first clearly recognized in early studies of neuron-target interactions (Oppenheim, 1981; Cowan, 2001), the role of targets in controlling PCD historically has received the most attention. However, as discussed in the previous section, signals derived from afferent inputs, as well as from nonneuronal cells such as central and peripheral glia and endocrine glands (e.g., steroid hormones), have now been shown to be possible sources of trophic regulation of cell death and survival (Fig. 18.13).

Studies of the regulation of vertebrate neuronal PCD have shown that targets are critically involved in regulating how many postmitotic cells in the innervating population survive or die; in this way the final number of surviving neurons can be maintained at an optimal level despite individual differences in target size. Complete or partial deletion of targets reduces survival, whereas increasing the size or number of available targets results in increased survival (Oppenheim, 1991).

As noted in a previous section, if too few neurons are generated by proliferation in individual animals, the extent of PCD in that population may be curtailed so as to preserve an optimal quantitative relationship between neurons and their synaptic targets (DeMarco Garcia & Jessell, 2008). Although it is thought that the relationship between neuronal survival and the availability of synaptic targets is proportional and linear (this has been called numerical-, quantitative-, size-, or systems-matching), there may also be other ways in which the number of neurons that innervate a target can be regulated. For example, signals from glial cells along axon pathways may modulate neuron numbers prior to target innervation. Spinal motoneurons provide one of the best examples of how systems-matching is regulated by a competition for target-derived signals (NTFs) that promote survival (the Neurotrophic Hypothesis).

However, before discussing the example of motoneurons, it is first necessary to discuss the principle of competition. Competition at the cellular level is a key principle in the conceptualization of neuronal death and survival. It occurs when individual cells with different properties interact with each other, directly or indirectly, and results in the survival of the "stronger" cell and the death of the "weaker" cell. However, despite its long history, neither the origins of nor the characteristics themselves that define weak versus strong neurons are well understood (Oppenheim, 1991).

Cells commonly adopt different phenotypes by virtue of either their lineage (autonomous specification) or cell–cell interactions (conditional specification). In some cases, however, cells may initially differ stochastically, independent of either lineage or cell–cell interactions, by the random generation of genetic noise and a means to amplify and stabilize the noise-generated differences between cells (Losick & Desplan, 2008). Although speculative, it seems plausible that the generation, amplification, and stabilization of genetic noise could provide a separate mechanism whereby neurons become "stronger" (winners) by virtue of, for example, their expression of NTF receptors or of other features that confer a competitive edge over "weaker" cells (the losers).

In the case of avian spinal motoneurons that innervate limb skeletal muscle, the number of neurons that survive the period of PCD bears a 1 : 1 relationship with the number of primary myotubes present in individual muscle precursors during the period of cell death, rather than being correlated with the final number of myotubes or myofibers (i.e., muscle size) present after the cessation of cell death. Accordingly, in this situation, motoneuron numbers are controlled, in part, by signals that are limited by the number of primary myotubes available. Although for many populations of neurons the essential factors provided by targets that mediate neuronal survival are not known, extrapolation from what is known for sensory, sympathetic, and motor neurons suggests that specific target-derived NTFs are involved. The PCD of CNS neurons in the avian isthmo-optic nucleus and neurons in the mammalian

thalamus and substantia nigra are controlled by many of the same mechanisms involved in the PCD of peripheral neurons, including a need for target- and afferent-derived signals. As discussed previously, the survival of myelin-forming glial cells also involves competition, in this case, competition for trophic signals derived from axons. In both neurons and glia, the final outcome of this competitive process (Figs. 18.3 and 18.4) is the survival of optimal numbers of cells for innervation (neurons) or myelination (glia).

Motoneurons (and some other neuronal populations as well, including retinal ganglion cells, isthmo-optic neurons, and ciliary ganglion neurons) have another interesting property in that their target dependency appears to be regulated by physiological synaptic interactions with their targets (Oppenheim, 1991). For example, following the formation of initial synaptic contacts between motoneurons and target muscles, the initiation of synaptic transmission activates the muscle and results in overt embryonic movements. Chronic blockade of this activity during the cell death period with specific drugs or toxins that cause paralysis, or in genetic mutants lacking synaptic transmission, prevents the death of all motoneurons (Fig. 18.14).

Because PCD is enhanced after the pharmacological blockade of afferent synaptic activity, the functional input provided by afferents also appears to be of fundamental importance in this situation. Functional afferent input may act to regulate the survival of both developing and adult-generated postsynaptic neurons by several different mechanisms:

1. Neurotransmitter release results in depolarization by afferent signals that can alter intracellular calcium levels in postsynaptic cells, which in turn can independently modulate survival.
2. Afferent activity can regulate the expression of NTFs and their receptors in postsynaptic cells.
3. The activity-dependent release of NTFs from terminals of afferent axons or from adjacent glial cells may also provide survival signals to postsynaptic cells.

In addition to the survival-promoting role of afferent-derived signaling, in some neuronal populations afferents may also have the counterintuitive role of promoting either death or survival depending, in part, on the extent to which receptor-mediated afferent signals change calcium influx: high influx promotes death, and low or moderate influx promotes survival. Because these neurons may also depend upon target-derived NTF signaling for survival, when taken together this kind of evidence indicates that the original version of the neurotrophic hypothesis is no longer sufficient for understanding the complexity of the cell–cell interactions involved in regulating the survival and death of developing neurons (Fig. 18.14). Competitive differences between neurons in a population may involve success in competing for survival-promoting signals as well as competitive success in avoiding death-promoting signals (Fig. 18.15) and these signals may include activity-independent signals derived from targets, afferents, and nonneuronal cells. In the final analysis, the various target- and afferent-derived extrinsic signals influence survival (or death) by modulating intracellular signaling pathways that regulate the expression of pro- and anti-apoptotic gene products (Figs. 18.7 and 18.12).

Although the cellular mechanisms are not yet well established, the survival and death of newly generated neurons in the adult hippocampus (Fig. 18.16) and olfactory system provide another example of how afferent activity can regulate these events. For example, learning, stress, motor activity, sensory stimuli, and other environmental influences appear to control granule cell numbers in the dentate gyrus of the hippocampus by activity-dependent modulation of neuronal survival, and odor experience modulates the survival of adult-generated olfactory neurons. Similar to developing neurons during embryonic and postnatal stages, the default state of developing adult-generated neurons is PCD and their survival, differentiation, and incorporation into functional circuits requires synaptic input and NTF signaling (Gage et al., 2008). Although there is still debate and controversy over whether and if so how adult-generated neurons contribute to brain functions that mediate learning, memory, and neural plasticity, increasing evidence supports the role of adult neurogenesis in specific behavioral adaptations.

Summary

During development, the fate of a cell, including the decision to live or die, can be determined by intrinsic cell-autonomous mechanisms or by extrinsic signals derived from cell–cell interactions. The survival of most developing neurons is likely to be dependent on a variety of signals, including multiple NTFs derived from diverse sources, that serve to maintain survival and regulate differentiation in complex ways that reflect the specific requirements of neurons at each step in their development. Targets and afferents provide critical survival-promoting signals that prevent PCD. One major class of such signals are NTFs. Neurons are thought to compete for limiting amounts of these NTFs. For many populations of neurons, synaptic transmission and physiological activity also plays an important role in regulating survival. Accordingly, both NTFs and electrical activity/synaptic transmission are important for the regulation of PCD and are closely linked functionally.

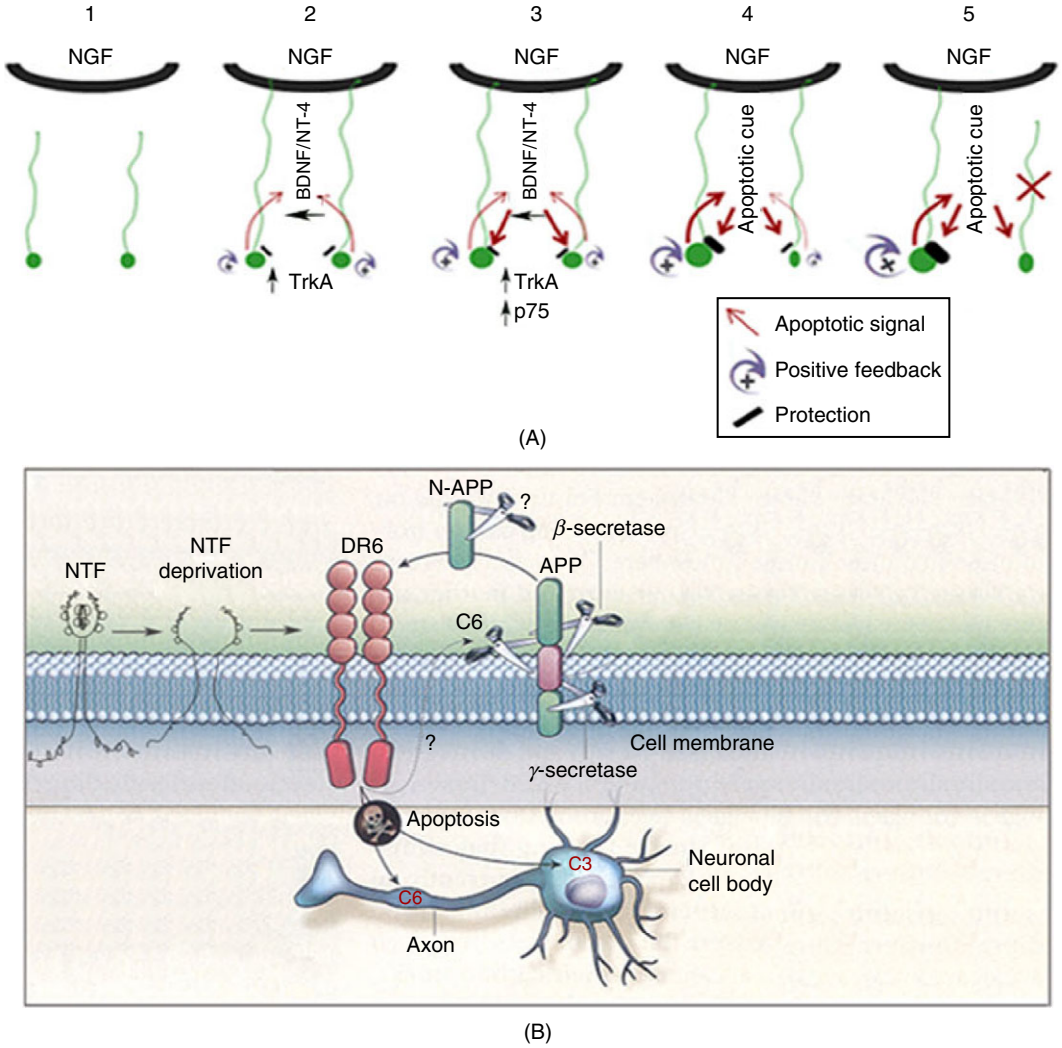


FIGURE 18.15 In (A) a model for developmental competition based on studies of developing sympathetic neurons is illustrated: (1) Before target innervation neurons (green) are modestly responsive to NGF; (2) upon target innervation and exposure to NGF, levels of TrkA, then BDNF and NT-4 (not shown) are increased; (3) induction of p75NTR expression, as well as differential sensitization of neurons, by modulation of NGF-TrkA signal strength and duration. (4) BDNF and NT-4 (apoptotic cues) kill neurons with low NGF-TrkA signaling; neurons with high NGF-TrkA signaling are resistant; (5) selection and neuronal death (*Adapted from Deppmann et al., 2008, Science* ***320****, 369-373*). In (B) a model of neurodegeneration involving the selective loss of developing neurons and their axons is illustrated. Following NTF deprivation the transmembrane amyloid-ß precursor protein (APP) is routinely cleaved by ß- and γ-secretase enzymes to generate both amyloid-ß peptide (Aß) and the amino-terminal portion of APP (N-APP). The soluble N-APP may then undergo further processing before binding to the death receptor DR6. DR6 responds to interaction with N-APP by engaging and activating key mediators of apoptotic cell death: caspase-3 (C3) in the neuronal cell body and caspase-6 (C6) in axons. Caspase-6 can also cleave APP near the ß-secretase target site, potentially contributing to the formation and/or amplification of this apoptotic circuit (*adapted from Nicholson, 2009, Nature* **457**, *970-971*).

THE ROLE OF TROPHIC FACTORS AND PROGRAMMED CELL DEATH IN NEUROPATHOLOGY

The role of trophic factors and the widespread occurrence of PCD during normal development indicate that cell loss, together with cell production (proliferation), is a fundamental mechanism for controlling final cell numbers in many tissues. During development trophic factors promote neuronal development, survival, and function. Therefore, it is not surprising that much research has been devoted to exploring their use as therapeutic agents to protect neurons in neuropathology (Lanni, Stanga, & Govoni, 2010). Trophic factor administration has been explored in animal models of many disorders of the nervous system, including the neurodegenerative diseases such as Parkinson's Disease (PD) and Amyotrophic Lateral Sclerosis (ALS) (Rangasamy, Soderstrom, Barkay, & Kordower, 2010; Gould & Oppenheim, 2011).

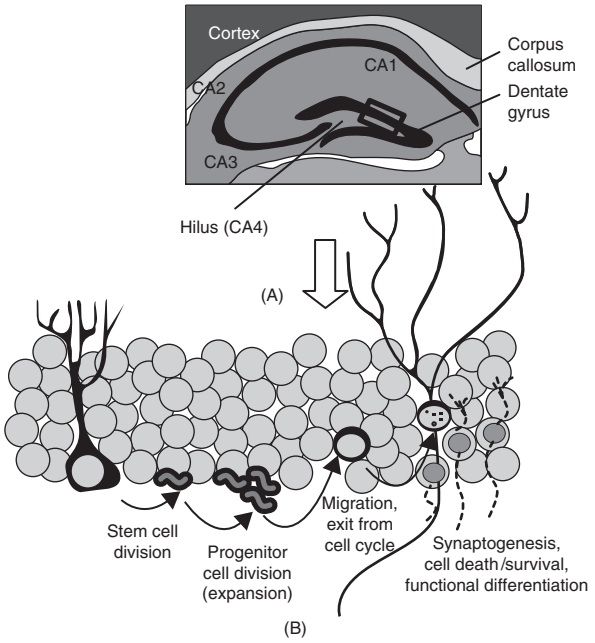


FIGURE 18.16 Schematic illustration of the hippocampus (A) including the dentate gyrus, a portion of which is shown enlarged in B. (B) The dentate gyrus is comprised of several layers of differentiated granule cells (light gray circles and dark cell with dendrites on left). Stem cells in lower subgranule layer of dentate gyrus undergo cell division and a subset of these migrate into the granule layer and differentiate (cell with dendrites on right) whereas others migrate and begin differentiation but then undergo PCD (small dark cells with dashed axons and dendrites). *Modified and redrawn from Kempermann (2006).*

In ALS, motoneurons become dysfunctional and die. Many trophic factors such as IGF-1, BDNF, GDNF, and CNTF that promote motoneuron survival during development have been administered to the ALS mouse model with positive results. These encouraging results of preclinical studies where administration of trophic factors promoted neuronal survival and delayed disease progression led to numerous clinical trials in ALS patients. Unfortunately, the results from the animal model studies did not translate into positive results in clinical trials. The failure of the clinical trials, however, does not warrant dismissal of the therapeutic potential of trophic factors, but rather brings to light the complexity of developing effective therapeutics for neurodegenerative diseases, including trial design, delivery mechanisms, dose and cellular targets.

The dysregulation of normal PCD can be maladaptive and pathological. Developmental, genetic, or congenital neurological defects may be caused by perturbations of PCD. For example, one proposed mechanism underlying autism is that there is a delay or absence of PCD (Courchesne et al., 2011). In the mature CNS, cell death exhibiting morphological characteristics and mechanistic components of PCD is a hallmark of many neuropathological conditions (Table 18.5). Neuropathology is a key example of why morphological characteristics of dying cells must be interpreted with caution. While DNA fragmentation and damage are observed in dying cells in pathological conditions, rarely are cells with morphological features of apoptosis observed; rather, dying cells appear to exhibit a combination of features characteristic of PCD types II, III, and necrotic cell death discussed above.

Autophagy is thought to be an adaptive response that can promote survival; however, dysregulation of autophagy is also proposed to contribute to neurodegenerative disorders (Banerjee, Beal, & Thomas, 2010). Although the morphological features may be different, many pathways of developmental cell death (see Fig. 18.7) can be re-activated in mature cells following insult or injury. As with developmental PCD, the mitochondria appear to have a central role in initiating cell death in epilepsy, Alzheimer's disease (AD), PD, ALS, and stroke (Martin, 2010). The Bcl-2 proteins that have key roles in mediating release of pro-death factors from the mitochondria have altered expression or interactions in neuropathologies. Caspase activation has also been observed in postmortem tissue from AD, PD, ALS, and stroke patients.

In some of these diseases, specific gene mutations have been identified, and the identification of these mutations has facilitated the development of transgenic or cellular models of these diseases. Studies of these model systems indicate that the mutated proteins may directly or indirectly perturb inter- and intracellular mechanisms involved in the maintenance and survival of specific populations of neurons. For example, mutations in PTEN-inducing putative kinase 1 (PINK1) and Parkin are associated with some familial forms of PD, and these mutated proteins may function to modulate mitochondrial degradation. Mutations in copper, zinc superoxide dismutase 1 (SOD1) are associated with some forms of familial ALS. Misfolded mutant SOD1 has been shown to associate specifically with Bcl-2 in mitochondria, and this SOD1-Bcl-2 interaction appears to be toxic. In familial AD, mutations in the amyloid precursor protein (APP) genes have been identified. APP, amyloid-β, and the cytoskeletal protein tau are major contributors to the plaques and tangles observed in AD brains. Although the secretases appear to be the major proteases that modulate cleavage of APP, APP has also been shown to be a substrate for caspases, and inhibition of caspase activity resulted in decreased production of Aβ, a molecule thought to mediate neurotoxicity in AD. Tau is also a substrate for caspases, and the caspase-cleaved tau fragment is a component of the neurofibrillary tangles.

The examples above illustrate potential roles for mutated proteins in pathological cell death; however, other mutations may affect how damaged or dying cells are cleared. For example, mutations in progranulin resulting in decreased expression of the protein are associated with frontotemporal dementia. Interestingly, in

TABLE 18.5 Neuropathological Conditions in which Aberrant Cell Death Is Implicated[a]

Developmental Disorders	Courchesne et al., (2007). Neuron. 56(2):399–413; Courchesne et al., (2011), JAMA 306:2001–2010
autism	
Down's syndrome	Sawa (2001). J. Neural Trans. Suppl. 61:361–71
medullobastomas and neuroblastomas	Johnson et al., (2009). Apoptosis 14:424–438
fetal alcohol syndrome	Cortese et al., (2006). Neurotoxicol Teratol 28:597–606
sudden infant death syndrome	Sparks and Hunsaker (2002). Childs Nerv Syst 18:568–592
spinal muscular atrophy	Lorson et al., (2010). Hum. Mol. Gen. 19:R111–8.
leukodystrophies	Southwood and Gow (2001). Microsc Res Tech 52:700–8
neonatal hypoxia-ischemia	Northington et al., (2011). Ann Neurol 69:743–758
Triplet Repeat Disorders	Timchenko and Caskey (1999). Cell Mol Life Sci 55:1432–1447
Huntington's Disease	
spinal bulbar muscular atrophy	
Machado-Joseph disease (SCA 3)	
spinocerebellar ataxia (SCA) 1, 2, 6	
fragile X syndrome	
Friedreich's ataxia	
myotonic dystrophy	
Psychiatric Disorders	Jarskog et al., (2005). Prog Neuropsychopharm Bio Psychiatry 29:846–858.
schizophrenia	Hu et al., (2011). Prog Neuropsychopharm Bio Psychiatry 35:11–7.
bipolar disorder	McKernan et al., (2009). Prog. Neurobiol. 88:246–63.
depression	
Trauma and Acquired Disorders	Liberski et al., (2008). Folia Neuropathol. 46:1–25.
prion diseases	Sabri et al., (2003). Brain Pathol. 13:84–94.
neuropathology of HIV infection	Artemiadis and Anagonostouli (2010). Eur. Neurol. 63:65–72.
multiple sclerosis	Krakstad and Chekenya (2010). Mol. Cancer 9:135-147.
gliobastomas	Friedman (2010). Neuroscientist 16:244-252/
epilepsy	Hasegawa et al., (2011). Acta Neurochir Suppl: 110:43-8.
subarachnoid hemorrhage	Hazell (2009). Neurochem Int 55:129-135
Wernicke's encephalopathy	Rowland et al., (2008). Neurosurg Focus 25:E2
spinal cord injury	Engel et al., (2011). J. Cerebral Blood Flow and Metabolism 31:1196–1210
traumatic brain injury and ischemia/stroke	
Neurodegenerative Diseases	Ribe et al., (2008). Biochem J 415:165–182
Alzheimer's	Traub et al., (2011). Curr Neurol Neurosci Rep. 11:67–77; Pedrini et al., (2010). Hum. Mol. Genet 19:2974–86
amyotrophic lateral sclerosis	
Parkinson's	Vives-Bauza and Przedborski (2011). Trends Mol Med 17:158–65
Other	
diabetic retinopathy	Barber et al., (2011). Invest Ophthalmol Vis Sci 52:1156–1163
glaucoma	Danesh-Meyer (2011). Curr Opin Opthalmol 22:78–86.
chronic pain	Apkarian and Scholz (2006). Ann Neurol 69:13–21.
Hirschsprung's disease	Uesaka et al., (2007). Development 134:2171–81.

[a]*Enhanced or reduced cell death has been associated with many developmental and neuropathological conditions listed in the table. Also provided are review articles or reports describing the aspects of cell death identified in each disorder.*

C. elegans mutants lacking progranulin, there are fewer apoptotic corpses observed, consistent with the suggestion that the clearance of dying cells may contribute to disease processes (Kao et al., 2011).

The idea that aberrant cell death in the developing and adult nervous system may reflect a loss of normal control mechanisms for survival is appealing in that the increased understanding of the genetic regulation of PCD provides a potentially powerful and rational approach for the development of therapeutic treatment strategies. However, recent evidence questions the validity of the widely held assumption that it is the death of neurons that is the primary cause rather than the result of the pathophysiology in these diseases. There is growing evidence that neuronal and synaptic dysfunction that precedes any significant neuronal loss may be a major factor in the functional and behavioral pathology that characterizes many neurodegenerative diseases (Gould & Oppenheim, 2007; Saxena & Caroni, 2011). Therefore, identification of the initiating factors leading to cell dysfunction and death will be critical for the development of effective therapeutics.

Summary

Because PCD is primarily a developmental phenomenon, historically the major focus of investigation has been the normal biology of cell death in the embryo, fetus, and newborn. However, with the growing recognition that pathological neuronal cell death may share certain biochemical and molecular features with PCD, there is growing hope that a better understanding of PCD may reveal potential therapeutic strategies for the treatment of developmental neuropathologies, neurodegenerative disease, and neuronal loss following CNS trauma and disease.

References

Airaksinen, M. S., & Saarma, M. (2002). The GDNF family: Signaling, biological functions and therapeutic value. *Nature Review Neuroscience, 3,* 383–394.

Ameisen, J. C. (2004). Looking at death at the core of life in the light of evolution. *Cell Death Differentiation, 11,* 4–10.

Banerjee, R., Beal, M. F., & Thomas, B. (2010). Autophagy in neurodegenerative disorders: pathogenic roles and therapeutic implications. *Trends in Neuroscience, 33,* 541–549.

Barde, Y. A. (1989). Trophic factors and neuronal survival. *Neuron, 2,* 1525–1534.

Barker, P. A. (2004). p75NTR is positively promiscuous: novel partners and new insights. *Neuron, 42,* 529–533.

Bredesen, D. E., Rao, R. V., & Mehlen, P. (2006). Cell death in the nervous system. *Nature, 443,* 796–802.

Buss, R. R., Sun, W., & Oppenheim, R. W. (2006). Adaptive roles of programmed cell death during nervous system development. *Annual Review of Neuroscience, 29,* 7–35.

Chao, M. V. (2003). Neurotrophins and their receptors: A convergence point for many signaling pathways. *Nature Review Neuroscience, 4,* 299–309.

Clarke, P. G. H., & Oppenheim, R. W. (1995). Neuron death in vertebrate development: In vivo methods. *Methods in Cell Biology, 46,* 277–323.

Clarke, P. G. H., & Clarke, S. (1996). Nineteenth century research on naturally occurring cell death, and related phenomena. *Anatomy and Embryology, 193,* 81–99.

Copani, A., Uberti, D., Sortino, M. A., Bruno, V., Nicoletti, F., & Memo, M. (2001). Activation of cell-cycle-associated proteins in neuronal death. *Trends in Neuroscience, 24,* 25–31.

Courchesne, E., Mouton, P. R., Calhoun, M. E., Semendeferi, K., Ahrens-Barbeau, C., Hallet, M. J., Barnes, C. C., & Pierce, K. (2011). Neuron number and size in prefrontal cortex of children with autism. *JAMA, 306,* 2001–2010.

Cowan, W. M. (2001). Viktor Hamburger, and Rita Levi-Montalcini: The path to the discovery of nerve growth factor. *Annual Review of Neuroscience, 24,* 551–600.

DeMarco Garcia, N. V., & Jessell, T. M. (2008). Early motor pool identity and muscle nerve trajectory defined by postmitotic restrictions in NKX6.1 activity. *Neuron, 57,* 217–231.

Deppmann, C. D., Mihalas, S., Sharma, N., Lonze, B. E., Niebur, E., & Ginty, D. D. (2008). A model for neuronal competition during development. *Science, 320,* 369–373.

Egan, M. F., Kojima, M., Callicott, J. H., Godlgerg, T. E., Kolachana, B. D., Bertolino, A., Zaitsev, E., Gold, B., Goldman, D., Dean, M., Lu, B., & Weinberger, D. R. (2003). The BDNF val66met polymorphism affects activity-dependent secretion of BDNF and human memory and hippocampal function. *Cell, 112,* 257–269.

Elliott, M. R., Chekeni, F. B., Trampont, P. C., Lazarowski, E. R., Kadl, A., Walk, S. F., Park, D., Woodson, R. I., Ostankovich, M., Sharma, P., Lysiak, J. J., Harden, T. K., Leitinger, N., & Ravichandran, K. S. (2009). Nucleotides released by apoptotic cells act as a find-me signal to promote phagocytic clearance. *Nature, 461,* 282–286.

Fuchs, Y., & Steller, H. (2011). Programmed cell death in animal development and disease. *Cell, 147,* 742–458.

Gage, F. H., Kempermann, G., & Song, H. (Eds.), (2008). *Adult neurogenesis.* Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Gauthier, L. R., Charrin, B. C., Borrell-Pages, M., Dompierre, J. P., Rangone, H., Cordelieres, F. P., DeMey, J., MacDonald, M. E., Lessmann, V., Huymbert, S., & Saudou, F. (2004). Huntingtin controls neurotrophic support and survival of neurons by enhancing BDNF vesicular transport along microtubules. *Cell, 118,* 127–138.

Gould, T. W., & Oppenheim, R. W. (2007). Synaptic dysfunction in disease and following injury in the developing and adult nervous system: Caveats in the choice of therapeutic intervention. *Neuroscience and Biobehavavioral Reviews, 31,* 1073–1087.

Gould, T. W., & Oppenheim, R. W. (2011). Motor neuron trophic factors: Therapeutic use in ALS? *Brain Research Reviews, 67,* 1–39.

Horvitz, H. R. (2003). Worms, life and death (Nobel lecture). *Chembiochem, 4,* 697–711.

Huang, E. J., & Reichardt, L. F. (2001). Neurotrophins: Roles in neuronal development and function. *Annual Review of Neuroscience, 24,* 677–736.

Ip, N., & Yancopoulos, G. (1996). The neurotrophins and CNTF: Two families of collaborative neurotrophic factors. *Annual Review of Neuroscience, 19,* 491–515.

Jiao, S., & Li, Z. (2011). Non-apoptotic function of Bad and Bax in long-term depression of synaptic transmission. *Neuron, 70,* 758–772.

Kao, A. W., Eisenhut, R. J., Martens, L. H., Nakamura, A., Haung, A., Bagley, J. A., Zhou, P., de Luis, A., Neukomm, L. J., Cabello, J., Farese, R. V., Jr., & Kenyon, C. (2011). A neurodegenerative disease mutation that accelerates the clearance of apoptotic cells. *Proceedings of the National Academy of Science of United States of America, 108,* 4441–4446.

Kempermann, G. (2006). *Adult neurogenesis: Stem cells and neuronal development in the adult brain.* New York: Oxford University Press.

Kinchen, J. M., & Ravichandran, K. S. (2007). Journey to the grave: signaling events regulating removal of apoptotic cells. *Journal of Cell Science, 120,* 2143–2149.

Kroemer, G., Galluzi, L., Vandenbeele, P., Abrams, J., Alnemri, E., Baehrecke, E., Blagosklonny, M., El-Deiry, W., Golstein, P., Green, D., Hengartner, M., Knight, R., Kumar, S., Malornd, W., Nunez, G., Peter, M., Tschopp, J., Yuan, J., Zhivotovsky, B., & Melino, G. (2009). Classification of cell death: recommendations of the nomenclature committee on cell death. *Cell Death and Differentiation, 16,* 3–11.

Lanni, C., Stanga, M., & Govoni, S. (2010). The expanding universe of neurotrophic factors: therapeutic potential in aging and age-associated disorders. *Current Pharmaceutical Design, 16,* 698–717.

Lauber, K., Bohn, E., Kröber, S. M., Xiao, Y. J., Blumenthal, S. G., Lindemann, R. K., Marini, P., Wiedig, C., Zobywalski, A., Baksh, S., Xu, Y., Autenrieth, I. B., Schulze-Osthoff, K., Belka, C., Stuhler, G., & Wesselborg, S. (2003). Apoptotic cells induce migration of phagocytes via caspase-3-mediated release of a lipid attraction signal. *Cell, 113,* 717–730.

Lessmann, V., Gottmann, K., & Malcangio, M. (2003). Neurotrophin secretion: Current facts and future prospects. *Progress in Neurobiology, 69,* 341–374.

Lewin, G. R., & Barde, Y. A. (1996). Physiology of the neurotrophins. *Annual Review of Neuroscience, 19,* 289–317.

Lindholm, P., & Saarma, M. (2010). Novel CDNF/MANF family of neurotrophic factors. *Development Neurobiology, 79,* 360–371.

Losick, R., & Desplan, C. (2008). Stochasticity and cell fate. *Science, 320,* 65–68.

Lu, B., Pang, P. T., & Woo, N. H. (2005). The yin and yang of neurotrophin action. *Nature Review Neuroscience, 6,* 603–614.

Mallat, M., Marín-Teva, J. L., & Chéret, C. (2005). Phagocytosis in the developing CNS: more than clearing the corpses. *Current Opinion in Neurobiology, 15,* 101–107.

Martin, L. J. (2010). Mitochondrial pathobiology in Parkinson's disease and amyotrophic lateral sclerosis. *Journal of Alzheimers Disease, 20*(Suppl 2), S335–356.

McAllister, A. K., Katz, L. C., & Lo, D. C. (1999). Neurotrophins and synaptic plasticity. *Annual Review Neuroscience, 22,* 295–318.

Mehlen, P. (2010). Dependence receptors: the trophic theory revisited. *Science Signaling, 3,* pe 47.

Nedelcu, A. M., Driscoll, W. W., Durand, P. M., Herron, M. D., & Raskidi, A. (2010). On the paradigm of altruistic suicide in the unicellular world. *Evolution, 65,* 3–20.

Nikoletopoulou, V., Lickert, H., Frade, J. M., Rencurel, C., Giallonardo, P., Zhang, L., Bibel, M., & Barde, Y. A. (2010). Neurotrophin receptors TrkA and TrkC cause neuronal death whereas TrkB does not. *Nature, 467,* 59–63.

Oppenheim, R. W. (1981). Neuronal cell death and some related regressive phenomena during neurogenesis. In W. M. Cowan (Ed.), *Studies in developmental neurobiology: Essays in honor of Viktor Hamburger* (pp. 74–133). New York: Oxford.

Oppenheim, R. W. (1991). Cell death during development of the nervous system. *Annual Review of Neuroscience, 14,* 453–501.

Poo, M. M. (2001). Neurotrophins as synaptic modulators. *Nature Review Neuroscience, 2,* 24–32.

Potts, R. R., Singh, S., Knezek, M., Thompson, C. B., & Deshinukh, M. (2003). Critical function of XIAP in regulating caspase activation during sympathetic neuronal apoptosis. *Journal of Cell Biology, 163,* 789–799.

Putcha, G. V., Deshmukh, M., & Johnson, E. M. (2000). Inhibition of apoptotic signalling cascades causes loss of trophic factor dependence during neuronal maturation. *Journal of Cell Biology, 149,* 1011–1017.

Rangasamy, S. B., Soderstrom, K., Barkay, R. A. E., & Kordower, J. H. (2010). Neurotrophic factor therapy for Parkinson's Disease. *Progress in Brain Research, 184,* 237–264.

Roux, P. P., & Barker, P. A. (2002). Neurotrophin signaling through the p75 neurotrophin receptor. *Progress in Neurobiology, 67,* 203–233.

Saxena, S., & Caroni, P. (2011). Selective neuronal vulnerability in neurodegenerative diseases: from stressor thresholds to degeneration. *Neuron, 71,* 35–48.

Schmitz, C., & Hof, P. R. (2005). Design-based stereology in neuroscience. *Neuroscience, 130,* 813–831.

Segal, R., & Greenberg, M. (1996). Intracellular signalling pathways activated by neurotrophic factors. *Annual Review of Neuroscience, 19,* 463–489.

Singh, K. K., Park, K. J., Hong, E. J., Kramer, B. M., Greenberg, M. E., Kaplan, D. R., & Miller, F. D. (2008). Developmental axon pruning mediated by BDNF-p75NTR-dependent axon degeneration. *Nature Neuroscience, 11,* 649–658.

Snider, W. D. (1994). Functions of the neurotrophins during nervous system development: What the knockouts are teaching us. *Cell, 77,* 627–638.

von Bartheld, C. S., & Fritzsch, B. (2006). Comparative analysis of neurotrophin receptors and ligands in vertebrate neurons: Tools for evolutionary stability or changes in neural circuits? *Brain Behavior and Evolution, 68,* 157–172.

von Bartheld, C. S., Wang, X., & Butowt, R. (2001). Anterograde axonal transport, transcytosis, and recycling of neurotrophic factors: The concept of trophic currencies in neural networks. *Molecular Neurobiology, 24,* 1–28.

Wyllie, A. H., Kerr, J. E., & Currie, A. R. (1980). Cell death: the significance of apoptosis. *International Review of Cytology, 68,* 251–306.

Zweifel, L. S., Kuruvilla, R., & Ginty, D. D. (2005). Functions and mechanisms of retrograde neurotrophin signaling. *Nature Review Neuroscience, 6,* 615–625.

Suggested Readings

Hamburger, V. (1992). History of the discovery of neuronal death in embryos. *Journal of Neurobiology, 23,* 1116–1123.

Huang, E. J., & Reichardt, L. F. (2001). Neurotrophins: Roles in neuronal development and function. *Annual Review of Neuroscience, 24,* 677–736.

Jacobson, M. (1991). *Developmental Neurobiology*. New York: Plenum.

Jellinger, K. A. (2010). Basic mechanisms of neurodegeneration: a critical update. *Journal of Cellular and Molecular Medicine, 14,* 457–487.

Levi-Montalcini, R. (1987). The nerve growth factor 35 years later. *Science, 237,* 1154–1162.

Oberst, A., Bender, C., & Green, D. R. (2008). Living with death: The evolution of the mitochondrial pathway of apoptosis in animals. *Cell Death and Differentiation, 15,* 1139–1146.

Oppenheim, R. W. (1996). The concept of uptake and retrograde transport of neurotrophic molecules during development: History and present status. *Neurochemical Research, 21,* 769–777.

Oppenheim, R. W., Caldero, J., Esquerda, J., & Gould, T. (2001). Target-independent programmed cell death in the developing nervous system. In A. F. Kalverboer & A. Gramsbergen (Eds.), *Brain, and behaviour in human development* (pp. 343–407). Dordrecht, The Netherlands: Kluwer.

Purves, D. (1988). *Body and brain: A trophic theory of neuronal connections*. Cambridge, MA: Harvard.

Putcha, G. V., & Johnson, E. M. (2004). Men are but worms: neuronal cell death in C. Elegans and vertebrates. *Cell Death and Diffierentiation, 11,* 38–48.

Ronald W. Oppenheim, Carolanne E. Milligan, and Christopher S. von Bartheld

This page intentionally left blank

CHAPTER

19

Synapse Elimination

OVERVIEW

In the development of both central and peripheral nervous systems, synapse formation generates some connections that exist only transiently. Proof that synapses are eliminated comes from both anatomical and physiological studies showing that in many parts of the mammalian nervous system, presynaptic axons are connected to postsynaptic partners during development that they are not connected to at later stages. One unambiguous example of such loss is the change in innervation of mammalian neuromuscular junctions in early postnatal life. At birth, each neuromuscular junction receives convergent innervation from multiple motor axons (Fig. 19.1A and 19.1C). In adult muscles, however, only a single motor axon innervates each neuromuscular junction (Fig. 19.1B). Also in the sympathetic and parasympathetic autonomic ganglia of the peripheral nervous system, axons transiently supply extra inputs to each ganglion cell (Fig. 19.2). In the developing mammalian central nervous system (CNS), there are also several cases of documented synapse elimination (Table 19.1). For example, multiple climbing fibers innervate each cerebellar Purkinje cell at birth, whereas all but one is removed in early postnatal life. In the auditory system, developing neurons in the magnocellularis nucleus receive innervation from several inputs but in later life only one axonal input per neuron remains. Perhaps the best-known examples of synapse elimination occur in the visual system where some neurons that are initially activated by input from both eyes become monocularly driven later on.

In all these cases, synapse elimination reduces the number of axons that innervate target cells, making the term *input elimination* perhaps a better descriptor. Input elimination, however, is only easy to measure in relatively unusual situations, such as when neurons end up with one axon or innervation from one eye. It remains unclear how common synapse elimination is in the more typical situation where neurons receive hundreds or even thousands of different axonal inputs from a diverse set of presynaptic cells. Thus in most parts of the developing nervous system, it remains unknown whether synapse elimination is occurring at all. The uncertainty is further confounded by contemporaneous synapse formation in the developing nervous system. As described following, some of the additional synapses are originating from axons that are already connected to a postsynaptic cell. Thus the *total* number of synapses impinging on a postsynaptic cell may even be increasing at the stage when there is a net reduction in the number of axonal inputs innervating the target cell. Thus assaying for synapse elimination requires quantifying a change in the number of innervating axons rather than counting the number of synapses, which in turn requires a means of counting axons, or at least having a situation where distinct but overlapping pathways can be separately labeled or stimulated (such as the two eyes). It seems likely that there is far more synapse elimination during development than presently documented. For example, stimulation of the optic nerve with progressively larger voltages to estimate the number of different retinal ganglion cells innervating individual lateral geniculate neurons has revealed a dramatic change from more than 20 innervating axons to only several during the first postnatal month (Chen & Regehr, 2000), a much larger change than had been previously suspected in that system.

Importantly, synapse elimination, despite causing a reduction in the number of sources of innervation to a target cell, does not lead to a net weakening of synaptic drive on target cells. In the example just mentioned, the retinogeniculate axons that persist undergo a greater than 50-fold increase in synaptic efficacy, which more than compensates for the loss of other axonal inputs to the same target cells. Analogous increases in synaptic efficacy of inputs that

© 2013 Elsevier Inc. All rights reserved.

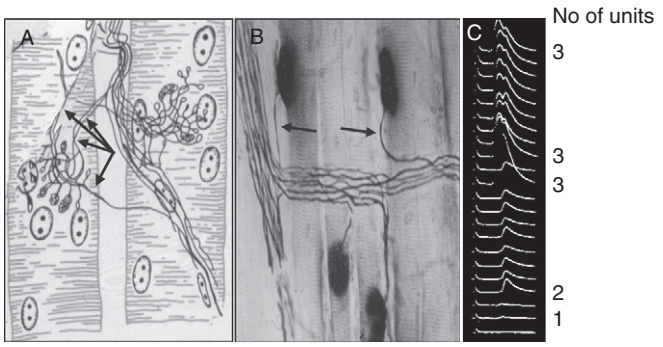


FIGURE 19.1 **Evidence of synapse elimination at the neuromuscular junction.** (A) Drawing of a silver stain made by J. Boeke in 1932 of the motor nerve terminals converging on two adjacent muscle fibers from the tongue muscle of a 4-day-old mouse. Each fiber is innervated by endings from several different axons (see arrows). (B) Photomicrograph of several adult muscle fibers from the rat extensor digitorum longus muscle. Each adult fiber receives innervation from a single axon (arrows). Axons are impregnated with silver and stained for cholinesterase (dark oval smudge on each muscle fiber). (C) Intracellular recordings from a muscle fiber in the neonatal rat diaphragm showing multiple innervation of a skeletal muscle fiber. Gradually increasing the strength of motor nerve stimulation elicits three different postsynaptic responses in this fiber. Similar assays in adult animals give only a single postsynaptic response that cannot be fractionated by graded stimulation. The preparation was partially curarized to prevent the endplate potentials from reaching threshold. *Figures adapted from (A) Boeke (1932); (B) Gorio, Marini, and Zanoni (1983); (C) Redfern (1970).*

persist have been described for the excitatory drive to neurons of the magnocellularis nucleus (Lu & Trussell, 2007) and for the climbing fiber inputs to Purkinje cells in the cerebellum (Hashimoto & Kano, 2003). In each case, the inputs that remain increase their efficacy by 10-fold or more as other axons are eliminated. The same trend was seen many years ago in the parasympathetic submandibular ganglion, where the remaining single input was always far more powerful than any of the inputs at the earlier stage when five to six multiple axons converged to the same single target cell (Lichtman, 1977). Similarly, at the neuromuscular junction, the remaining motor axon increases its quantal content as other inputs are eliminated (Colman, Nabekura, & Lichtman, 1997). It thus appears that although some inputs are being removed, the survivors are potentiated. Such results raise the outstanding question of the degree to which the compensation is explained by changes in the efficacy of the maintained synapses or rather the elaboration of new synapses.

A partial answer to this question comes from studies in the peripheral submandibular ganglion. Here, the simplicity of the neuropil (most synapses are axosomatic) allows for direct counts of synapse number during the period of synapse elimination. In this ganglion, input elimination was accompanied by concurrent synapse addition by the

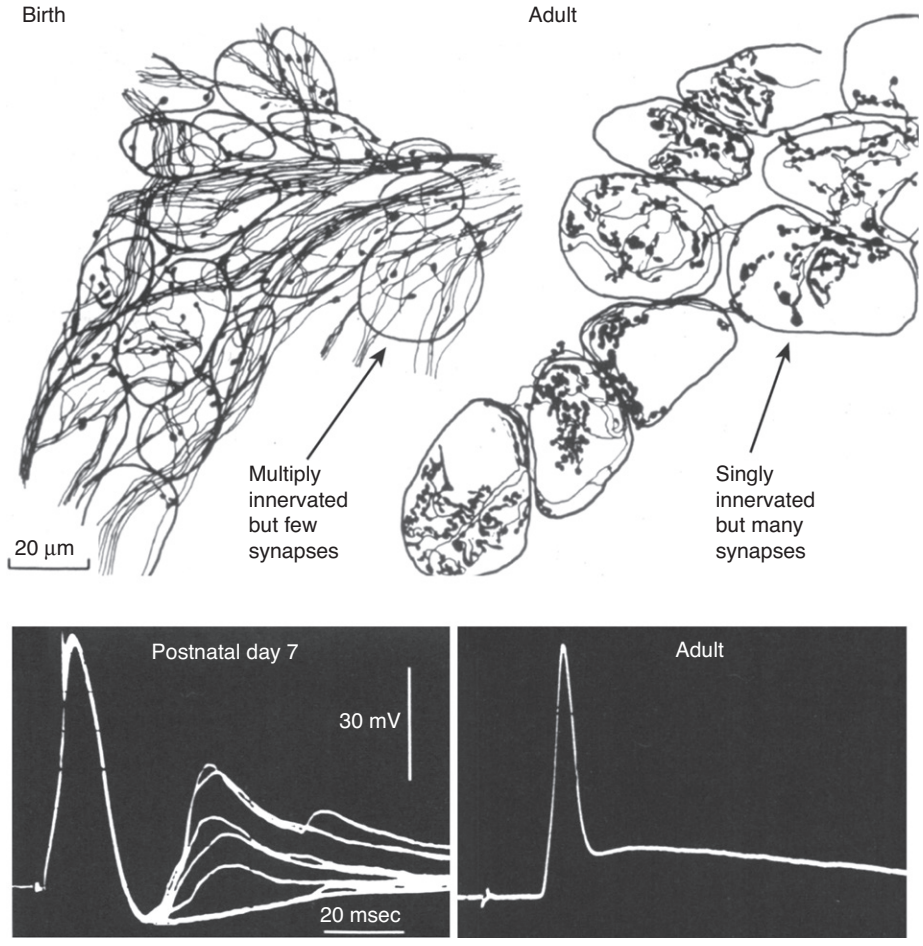


FIGURE 19.2 **In the developing nervous system synapse number is increasing while axonal convergence is falling.** The total number of preganglionic synapses on ganglion neurons in the rat submandibular ganglion increases in early postnatal life during the period of synapse elimination. Camera lucida drawings of clusters of ganglion neurons at birth and in adult animals that were treated with zinc iodide osmium. This reagent selectively fills terminal axons and synaptic boutons with a dense black precipitate. The number of synaptic boutons increases (compare the amount of synaptic boutons (arrows) at birth and on adult ganglion cells), and the number of preganglionic axons innervating each ganglion cell, identified by electrophysiological measurements, decreases (see the number of electrophysiological steps (arrow), each representing an input, at early postnatal life and in adult ganglion neurons, see lower panels). Thus axons that are not removed create new synapses to more than compensate for the loss of synapses from other axons. *From Lichtman (1977).*

TABLE 19.1 Synapse Elimination in the Mammalian Nervous System

Visual cortex: Layer IV	Binocular to monocular	Postnatal monkey, cat, ferret	(Hubel et al., 1977; Hubel & Wiesel, 1963; Wiesel, 1982)
Thalamus: Lateral geniculate nucleus	Binocular to monocular >20 to ~1–2 retinal axons	Postnatal cat and rat	(Chen & Regehr, 2000; Shatz, 1990)
Retina: Retinal ganglion cells	On/off to on- or off-center receptive fields	Postnatal ferret	(Wang, Liets, and Chalupa, 2001)
Cerebellum: Purkinje cell	>3 to 1 climbing fibers	Postnatal rat	(Mariani, 1983)
Cochlea: Magnocellularis cells	>4 to 1 cochlear axon	Chick	(Jackson & Parks, 1982; Lu & Trussell, 2007)
Parasympathetic: Submandibular ganglion	~5 to ~1 preganglionic axons	Postnatal rat	(Lichtman, 1977)
Sympathetic: Superior cervical ganglion	~14 to ~7 preganglionic axons	Postnatal hamster	(Lichtman & Purves, 1980)
Neuromuscular junction	2–6 to 1 motor axons	Postnatal rat, mouse	(Brown et al., 1976; Redfern, 1970)

remaining input. In particular, over the first postnatal month, there was approximately a twofold increase in the total number of synapses per ganglion cell, whereas the number of innervating axons was reduced fivefold (Fig. 19.2). This compensation indicates that the remaining axon on each ganglion cell must be increasing the number of synaptic contacts by approximately a factor of 10 as other axons lose all their connections to the same neuron. Such compensatory synaptogenesis means that assays of the total *number* of synapses at various developmental stages or following learning paradigms may belie a rather dramatic change in the *source* of the synapses.

It is also clear that in many situations the alterations occurring during this developmental reorganization are better described as a *redistribution of synapses* than as loss *per se*. The view is based on the fact that counts of the total number of neurons that innervate the entire target are not changing during the stage when the number of axons innervating each postsynaptic cell is decreasing (see, for example, Brown, Jansen, & Van Essen, 1976). Moreover, in several regions of the nervous system, it is known that synapse elimination occurs after the period of *naturally occurring cell death* is complete (see Chapter 18). Thus, as the elimination process reduces the number of axons converging on postsynaptic cells, it does not entirely remove axons from the postsynaptic target region but rather reduces the number of target cells innervated by each axon—that is, it reduces the axonal divergence (sometimes called fan-out) as individual axons restrict their synapses to a smaller number of postsynaptic target cells (Fig. 19.3). In muscle, axonal branch trimming during development has been long appreciated by an electrophysiological assay: the proportion of the total twitch tension an individual axon exerts decreases gradually over early postnatal life (Brown et al., 1976). This assay suggests that motor units (the term given to the number of muscle fibers innervated by a single motor axon) reduce their sizes by about threefold after birth.

Newer techniques of visualizing all the branches of single axons by transgenic expression of GFP (Feng et al., 2000) have, however, suggested that the changes are far more dramatic. Transgenic animals, in which only one axon is labeled, show that axonal arbors are about ten times larger at birth than in adults (Fig. 19.4A). These very large motor units imply that individual muscle fibers are innervated by ten or so axons because the total number of muscle fibers and neuromuscular junctions is fixed from birth onward. Direct evidence of this massive convergence on neuromuscular junctions was obtained by serial electron microscopy (Fig. 19.4B). Thus in muscle there is simultaneously extensive fan-in on neuromuscular junctions and fan-out of motor units in early life that is dramatically removed in early postnatal life.

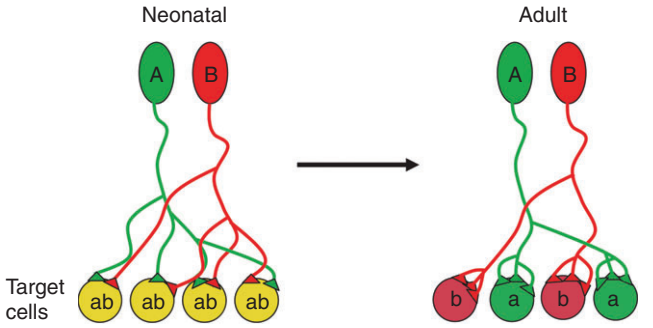


FIGURE 19.3 **Changes in fan-in and fan-out at developing circuits.** In neonatal vertebrates, neurons (A, green and B, red) project their axons to fan out to many target cells (each labeled ab). These nascent connections are typically weak and mediated only by a small number of synaptic contacts (triangles from the green and red axons). Axons undergo several structural rearrangements before reaching adulthood. First, axons disconnect from many target cells reducing their fan-out or divergence. Second, this rearrangement leads to less fan-in or convergence. Third, at the same time as axons disconnect from some target cells they are strengthening their connection with other target cells by increasing the number of synapses at their remaining targets.

The transgenic expression of fluorescent proteins has allowed study of how this removal occurs. Direct observation of axonal *branch retraction* in living animals by time-lapse imaging shows that axons that are removed first become thin (axonal "atrophy"); they then disconnect from the postsynaptic site and temporarily possess

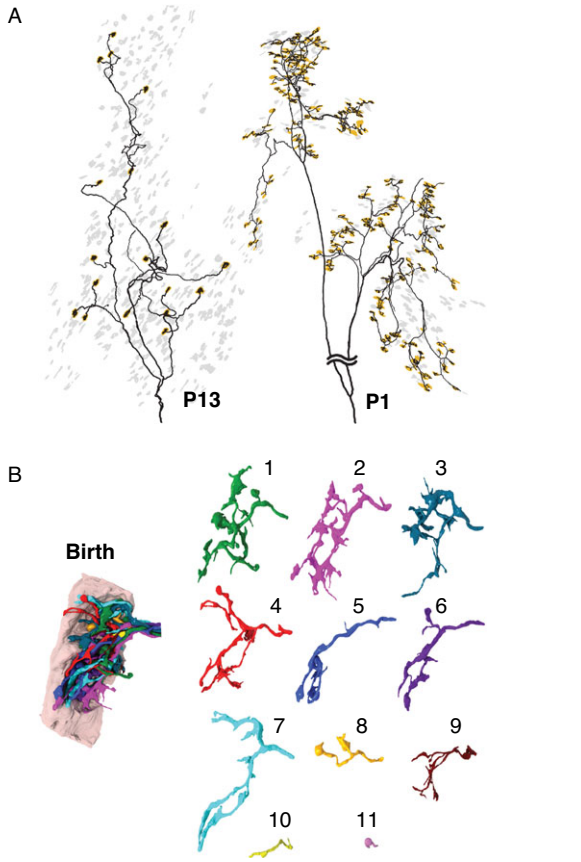


FIGURE 19.4 **Profound changes in axonal arbors and synapses at birth.** (A) Left, The axonal arbor of a motor axon's motor unit from a postnatal day 13 mouse. The neuromuscular junctions innervated by this axon are highlighted (yellow), and the rest of the neuromuscular junctions are shown in gray. Right, The branching arbor of a motor unit in the same muscle at postnatal day 1. At this young age the axon arbor is dramatically more complicated, and it innervates manyfold more neuromuscular junctions (yellow). B, The image on the left shows a serial electron microscopy 3D reconstruction of a single neuromuscular junction at birth that was then traced to show that 11 different axons converged on the site. On the right, each of the 11 axons that innervate this neuromuscular junction at birth are separately displayed, showing that they are quite intermingled and none occupies a majority of the neuromuscular junction.

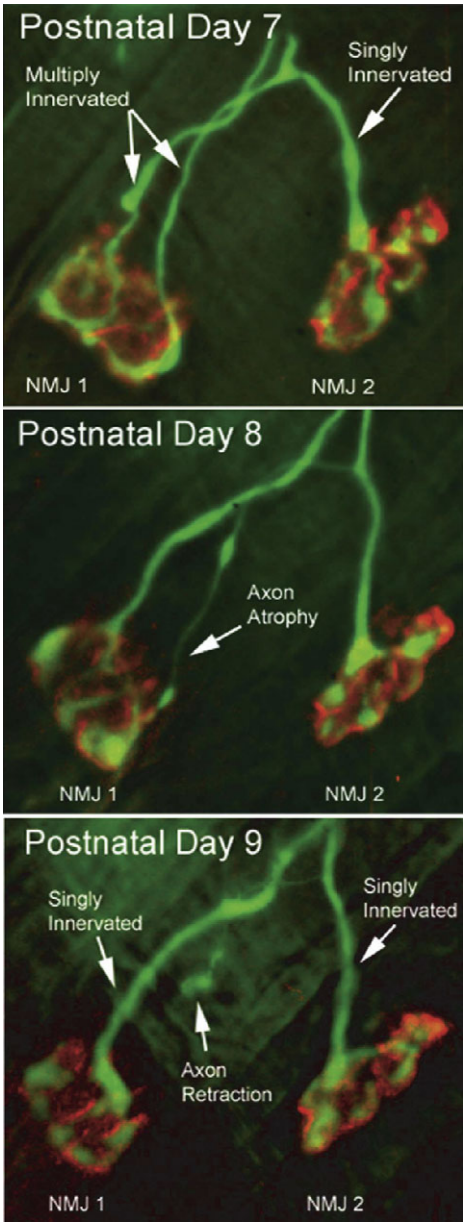


FIGURE 19.5 **Time-lapse imaging shows how axonal branches are lost during synapse elimination at the neuromuscular junction.** Two neuromuscular junctions (NMJ1 and NMJ2) were viewed *in vivo* on postnatal days 7, 8, and 9 in a transgenic mouse that expresses YFP in its motor axons (green). The acetylcholine receptors at the muscle fiber membrane are labeled with rhodamine-tagged α-bungarotoxin (red). On day 7, NMJ1 is multiply innervated, whereas NMJ2 is singly innervated by a branch of one of the axons that innervates NMJ1. One day later (postnatal day 8), one of the motor axons innervating NMJ1 becomes thinner all the way back to its branch point. When viewed on postnatal day 9, this thin branch is no longer connected to NMJ1 and now ends in a bulb-shaped swelling that is called a retraction bulb. *Modified from Keller-Peck et al. (2001).*

a bulb-tipped free axonal ending called a *retraction bulb* (Fig. 19.5). The withdrawing axon sheds small spherical membrane-enclosed axonal fragments filled with synaptic organelles (axosomes) to be engulfed by surrounding Schwann cells (Fig. 19.6). The sites formerly occupied by the withdrawing axon are mostly taken over by the axon that remains at the junction (described in detail following).

Summary

In the developing nervous system, both the number of axons that converge on postsynaptic cells and the number of postsynaptic cells an axon contacts decrease. The removal of inputs is due to a process of branch trimming. Concurrently, the remaining inputs appear to compensate by adding synaptic strength at least in part by the

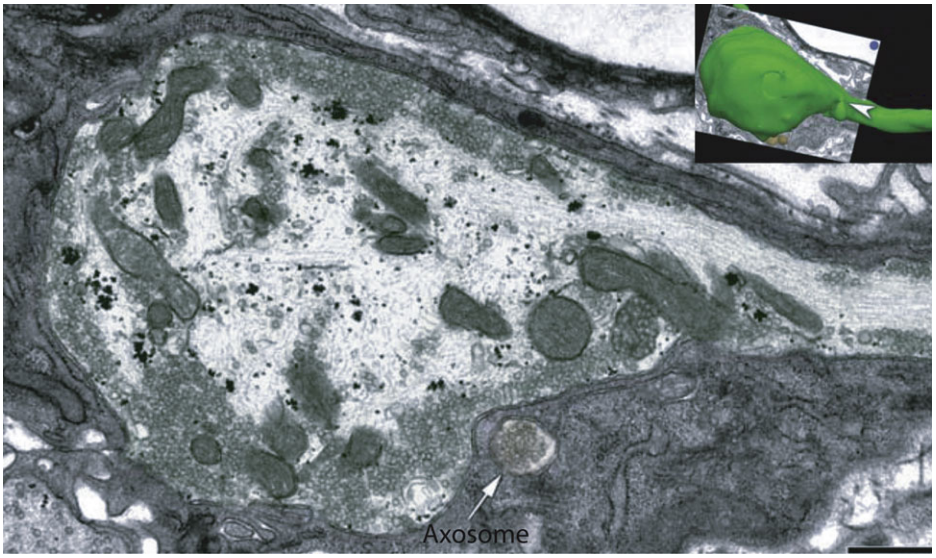


FIGURE 19.6 **Retracting axons shed material as they disappear.** The inset shows a surface rendering of a serial electron microscopy reconstruction of a retraction bulb (green, arrowhead) during synapse elimination at a developing neuromuscular junction. The electron micrograph shows that the retracting axon contains clusters of vesicles and mitochondria. In the vicinity of the bulb but not connected are small spherical axonal fragments (axosomes, arrow) that are engulfed by glial cells (i.e., Schwann cells (SC), darker cytoplasm). Scale bars, 1 mm. *From Bishop, Misgeld, Walsh, Gan, and Lichtman (2004).*

formation of new synapses. This redistribution refines synaptic circuitry by allowing an axon to strongly focus its innervation on a subset of the cells it initially contacted while each postsynaptic cell is restricted to responding to a subset of the axons that initially innervated it.

THE PURPOSE OF SYNAPSE ELIMINATION

At first sight, it might seem reasonable that synapse elimination is some form of error correction to rid the developing nervous system of outright connectivity mistakes. Indeed, there are clear examples of classes of nerve cell connections that disappear during development. For example, there are connections between mossy fiber axons and Purkinje cells in the developing cerebellum but not in the adult (Kalinovsky et al., 2011). In other situations, synapse elimination may *sharpen specificity* of the connections in a slightly more subtle way. For example, in some muscles there is a topographic map that matches the position of an axonal arbor to the location of the cell body within the motor neuron pool in the spinal cord. In these muscles synapse elimination makes these maps more precise (see, for example, Laskowski, Colman, Nelson, & Lichtman, 1998).

However, synapse elimination also occurs robustly in muscles and autonomic ganglia in which there is no evidence for intrinsic positional or other qualitative differences between axons that are maintained and those that are lost from any particular postsynaptic cell. Thus, while synapse elimination may be employed to refine the specificity of synaptic circuits, it seems to occur even in situations where it is not obvious that the synaptic connections that are maintained are from a different class of neuron than the ones that are eliminated.

An alternative hypothesis to the idea of error correction is that the loss of connections is a natural consequence of the extreme degree to which mammalian (and other vertebrate) nervous systems are composed of *duplicated neurons* (Lichtman & Colman, 2000). For example, pools of motor neurons that may number in the hundreds innervate individual skeletal muscles that contain thousands of muscle fibers. This reduplicated arrangement differs from the cellular plan in most parts of invertebrate nervous systems. For example, in the fly neuromuscular system, sometimes a single identified motor neuron innervates a muscle consisting of only one muscle fiber (Fig. 19.7A). Thus in mammals a cadre of nearly identical motor neurons that seem to have approximately equivalent roles project to a target consisting of a large population of roughly equivalent postsynaptic cells that constitute a muscle (Fig. 19.7B).

There are two potential circuit consequences of such redundancy in pre- and postsynaptic populations. First, many postsynaptic neurons are appropriate matches for each presynaptic axon, causing substantial axonal divergence or *fan-out* (Fig. 19.7C). Second, many presynaptic neurons could be appropriate matches for each postsynaptic cell, causing substantial axonal convergence or *fan-in* (Fig. 19.7D). Therefore, given the redundancy in both pre- and postsynaptic populations, one might expect overlapping *converging and diverging pathways* (Fig. 19.7E). Indeed this *is* the pattern of innervation in the neuromuscular systems in mammals and many other vertebrates, but only during development.

Therefore, what is surprising is that this overlap is short-lived and circuits undergo a radical transformation in postnatal life. As already mentioned, during neuromuscular development, axonal branches are pruned in a highly selective way, causing the projection

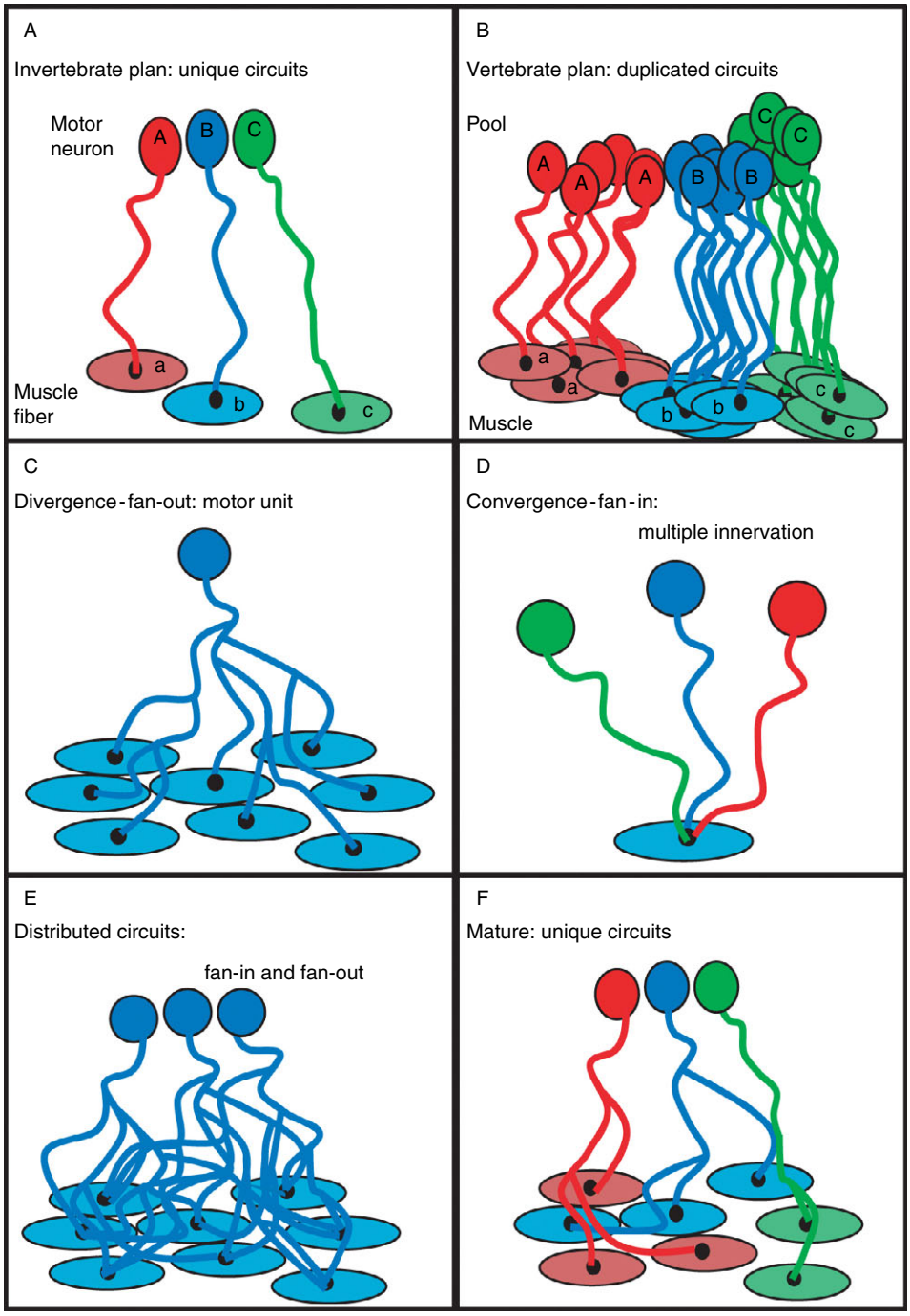


FIGURE 19.7 A diagram showing differences between the vertebrate and invertebrate synaptic circuits. (A) Invertebrates have small numbers (for example, sometimes just one motor neuron) of identifiable neurons innervating small numbers of identifiable target cells (e.g., sometimes just one muscle fiber), whereas in vertebrates, pools of similar neurons innervating targets contain hundreds or thousands of similar postsynaptic cells (B). The redundancy in the vertebrate nervous system allows a neuron to diverge (fan out) and innervate many equally appropriate target cells (in the neuromuscular system, this is called a motor unit) (C). The homogeneity of innervating neurons allows multiple axons to converge (fan in) on each target cell (D). As a result, vertebrate circuits contain a substantial amount of fan-in and fan-out (E), at least initially. Synapse elimination's purpose may be to transform such a set of redundant circuits into multiple unique ones by trimming away the multiple innervation of target cells (F). The result of the widespread loss of synapses is the generation of thousands of nonredundant circuits from an initially much less specific innervation pattern. *From Lichtman and Colman (2000).*

of each motor axon to become completely nonoverlapping and thus completely distinct from other axons' projections (Fig. 19.7F). From a functional standpoint, once synapse elimination is complete, the recruitment of each motor axon gives rise to activation of a different set of muscle fibers and consequently a substantive increase in tension of the muscle. This stepwise tension increase is a necessary aspect of the *size principle* (see Chapter 18). Therefore, the circuit rearrangement has an important consequence on neuromuscular output—that is, behavior.

By parsing highly redundant circuitry into multiple, unique, functionally distinct circuits, synapse elimination may be an essential maturation step for synaptic circuits that begin with considerable redundancy, such as those in our own nervous systems and those of other terrestrial vertebrates. It is possible that this paring down strategy is less relevant in other kinds

of animals. Indeed, in most invertebrates, identified neurons are the rule, and there is scant evidence for the type of neuronal redundancy seen throughout the mammalian nervous system. This difference suggests that a distinct neurodevelopmental strategy is at play in mammals and perhaps other vertebrate classes compared to the strategy used in the vast number of species in all the other subphyla. It is perhaps not a coincidence that another difference that partitions humans and other mammals from most invertebrates is that they seem highly dependent on experience for the acquisition of their behavioral repertoire. Invertebrate behavior is, to a greater degree, intrinsic and stereotyped. Compare, for example, the ease with which a newly hatched dragonfly takes wing versus the protracted period necessary for a human child to learn to walk. It is possible that the neuronal redundancy found in higher vertebrates that gives rise to overlapping convergent and divergent pathways is used in the acquisition of skills by the experience-mediated *selection* of connections.

This difference between mammals and invertebrates does not mean that synapses cannot be eliminated in invertebrates. Indeed, many studies have made the point that synaptic remodeling through addition and removal occurs in insect nervous systems (see, for example, Ding, Chao, Wang, & Shen, 2007; Eaton & Davis, 2003). The molecular mechanisms involved in invertebrate synaptic remodeling may provide insights into the mechanisms of synaptic loss and maintenance in higher vertebrate nervous systems. In worms, for example, the ubiquitin proteosome complex has been shown to regulate the elimination of motor neuron synapses onto vulval muscles (Ding et al., 2007). Similar degradation mechanisms may also be involved in synaptic remodeling of vertebrate neurons (Bingol et al., 2010; Li et al., 2010). The important point here, however, is that these mechanisms are used in mammals to alter the connectivity of the nervous system (i.e., to alter the number and identity of an axon's postsynaptic targets). In most invertebrate systems, and perhaps lower vertebrates such as fish, these same mechanisms may play more of a role in modifying the strength of existing connections without altering the circuit's wiring diagram.

Summary

In contrast to invertebrates, the nervous systems of terrestrial vertebrates contain reduplicated populations of neurons that serve each function. Elimination of synaptic connections during development may be an adaptation that converts highly overlapping connections of redundant neurons into unique circuits. Because this conversion may be based on experiences that affect the development of the nervous system, this process may tune the nervous systems of higher animals to the particular environment of each individual animal.

A STRUCTURAL ANALYSIS OF SYNAPSE ELIMINATION AT THE NEUROMUSCULAR JUNCTION

Thanks in large part to the power of fluorescence microscopy and the accessibility of the neuromuscular junction, it has been possible to describe the physical changes in axons and synapses that take place during synapse elimination. Interestingly, this purely descriptive analysis has yielded important clues and some mechanistic insights into the process of synapse elimination. As Yogi Berra said, "You can observe a lot by just watching." Imaging neuromuscular junctions at various ages in early postnatal life makes clear the fact that the process of synapse elimination is not a sudden calamitous event but rather a protracted process. At birth, multiple axons converging at a neuromuscular junction are highly intermingled, and the extent of the receptor areas occupied by each is similar (Balice-Gordon, Chua, Nelson, & Lichtman 1993). These highly intermingled connections eventually become progressively segregated over several days (Figs. 19.4B and 19.8). The partitioning of synaptic areas associated with different axons suggests that there is a spatial component to the mechanism. One scenario, for example, is that each axon *locally destabilizes* other inputs in their immediate vicinity but cannot as efficiently affect slightly more distant branches of the same axons. If the destabilization of an input was followed in turn by the takeover of its synaptic sites by the remaining axon, then an axon with a larger consolidated area might begin to destabilize more distant inputs. The idea of progressive consolidation was put to the test by the use of transgenic mice expressing different-color fluorescent proteins in individual axons. With these animals it is possible to obtain a precise map of the territories occupied by two axons at the same neuromuscular junction. *In vivo* time-lapse imaging of a multiple innervated neuromuscular junction over several days showed both gradual withdrawal of one axon from postsynaptic sites and a corresponding expansion of another axon to take over those synaptic sites (Fig. 19.9). This result may mean that axons are vying to occupy the same sites. Interestingly, it was not inevitable that the withdrawal of one axon was followed by the takeover of its sites by another input. In some cases, an axon that was already somewhat segregated from the other input would vacate its postsynaptic territory, but rather than being replaced by the remaining axon, its acetylcholine receptor-rich postsynaptic site would disappear. Synaptic loss without reoccupation by another input suggested that the takeover process *per se* was not causing withdrawal of the other axon. These observations suggest rather that synaptic takeover might be a response to the recent synaptic vacancy.

These time-lapse studies also demonstrated that the shift in favor of one axon was not irreversible. In some

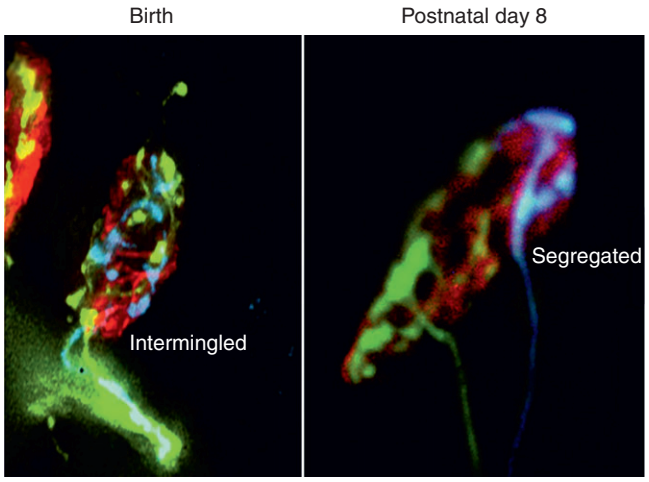


FIGURE 19.8 **Segregation of synaptic territory as axons compete during synapse elimination.** Double transgenic mice expressing YFP and CFP in different motor axons were used to observe axons' nerve terminal interactions. At birth is shown two competing nerve terminals, one expressing CFP (blue) and the other YFP (green) at this neuromuscular junction, which are intermingled. By the second postnatal week, however, YFP and CFP axon terminals are generally completely segregated from each other in multiply innervated junctions. Red is fluorescently tagged α-bungarotoxin that binds to AChRs. *Modified from Gan and Lichtman (1998).*

junctions, for example, it was hard to predict which axon might ultimately be maintained because the input with the majority of the territory shifted back and forth. Such flips in which axon seemed to be dominant suggest that the outcome is not preordained. Instead, axons at multiply innervated neuromuscular junctions appear to be in the midst of a highly dynamic interaction to determine which axon is most likely to stay.

Insights into the regulation of this dynamism come from analyses of each of the branches of individual motor axons. In lines of transgenic mice in which fluorescent proteins are expressed in a very small subset of motor neurons, it has been possible to examine all the branches of one axon during the developmental period when branches are being pruned. These studies show that axonal branch loss is occurring asynchronously among all the branches of one axon (Fig. 19.10). Thus, while some branches are recently eliminated (i.e., retraction bulbs), other branches are still connected to neuromuscular junctions that are multiply innervated. This range suggests that the fate of each branch is controlled independently. If axonal branches of one neuron are interacting with different axons at each of its neuromuscular junctions, then perhaps the rate of synapse

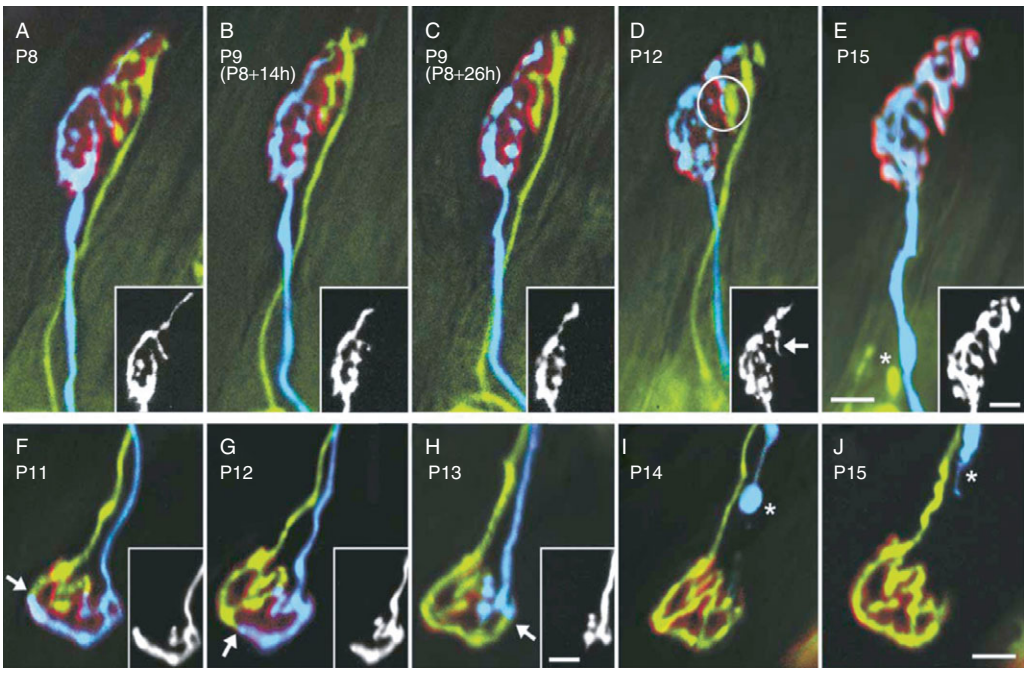


FIGURE 19.9 ***In vivo* imaging shows takeover of synaptic territory by the remaining axon during the period of synapse elimination.** Neuromuscular junctions in transgenic mice that express YFP (yellow) and CFP (blue) were imaged multiple times during early postnatal life. (A–E) views of one neuromuscular junction between P8 and P15. The CFP axon takes over occupancy of the postsynaptic sites (labeled red) in the upper parts of the junction that were formerly innervated by the YFP axon. The YFP labeled axon withdrew until only a retraction bulb remained (E, asterisk). At P12, a process of the CFP axon had begun to invade the territory of the YFP axon (D, circle and arrow in inset). (F–J) Although the CFP axon (blue and insets) has greater terminal area (~70%) at the first view, it progressively withdraws from the junction (arrows). Its retraction bulb can be seen in (I) and (J) (asterisks). Scale bars = 10 μm. Insets show the blue axons. *From Walsh and Lichtman (2003).*

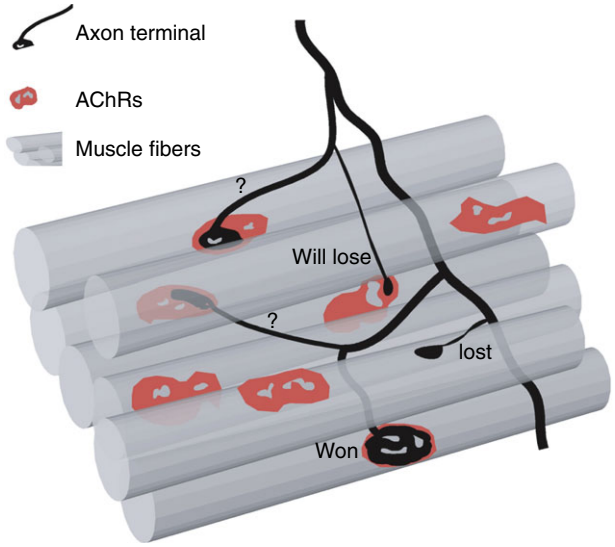


FIGURE 19.10 **Asynchronous synapse elimination among the branches of one axon.** Using transgenic animals that express YFP in a small subset of motor axons, it was possible to monitor the behavior of multiple branches of the same axon. This diagram shows the typical result for an axon in the midst of the synapse elimination process. Represented in red are the AChRs on each muscle fiber (represented as gray tubes). This motor axon (black) has won the competition on the bottom muscle fiber and occupies the entire receptor plaque. The same axon has lost the competition for the adjacent muscle fiber, where only a retraction bulb remains. The three other neuromuscular junctions that this axon innervates are still undergoing competition. In one case a small portion of AChRs is being innervated by this axon. It is likely that this synapse will be eliminated. On the other two muscle fibers, each axon terminal occupies 50% of the AChRs. This indicates that its fate is not yet determined in these junctions. *Adapted from Lichtman and Colman (2000).*

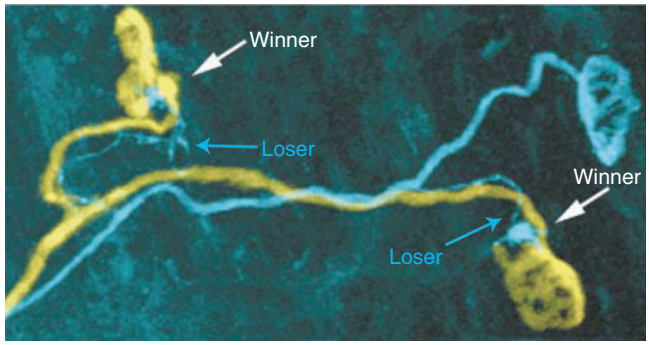


FIGURE 19.11 **Axonal fate at individual neuromuscular junctions is related to the identity of the competing axons.** Using transgenic animals that express YFP (yellow) and CFP (blue) in a small subset of motor axons, it was possible to reconstruct two motor units in which several neuromuscular junctions are innervated by the exact same two axons, one expressing YFP and the other CFP. At these coinnervated junctions, it appears that the same outcome is occurring at each. In this case, the YFP axon always occupied more territory than the CFP axon. Note also that the axon calibers of the YFP axon branches are thicker than CFP branches to the same junctions. Thus synapse elimination seems to be biased such that when the same two axons compete, the fate is the same at each coinnervated neuromuscular junction. *Image from Kasthuri and Lichtman (2003).*

elimination is regulated by which particular axons are co-innervating each of these junctions. These interactions might not only determine the rate but also the fates of the terminals—that is, which neuron's branches are maintained and which are eliminated. For example, an axon might be eliminated quickly at neuromuscular junctions innervated by some other neurons but fare better at junctions innervated by other axons.

To explore this possibility, transgenic mice with only two labeled axons (one yellow and one cyan) were generated to examine each of the neuromuscular junctions co-occupied by the same two axons within a developing muscle. The result was dramatic: at each of the shared neuromuscular junctions, the two axons often seemed to be in the same relative state (Fig. 19.11). Thus if the cyan axon was occupying only a small amount of territory at one junction that it shared with the yellow axon, then it occupied a small amount of territory at all or nearly all the other junctions it shared with the same yellow axon. However, where the cyan axon was interacting with other axons, its fate could be quite different. This result argued that the fate of axonal branches and the rate at which they are eliminated are related to the identity of the co-innervating axons. But why might the yellow axon be consistently "better" than the blue at all the co-innervated neuromuscular junctions they share? Counts of the total number of neuromuscular junctions each axon innervated (i.e., its motor unit size) provided a hint; the outcome of synapse elimination seemed to depend on the relative sizes of the two axons' motor units. In particular, neurons with larger motor units (such as the cyan axon in Fig. 19.11) were at a disadvantage when confronting neurons with smaller arborizations (such as the yellow axon in Fig. 19.11). One interpretation of this result is that axons with few branches in a muscle could dedicate more resources to a multiply innervated neuromuscular junction than neurons with a larger number of branches that are in some sense overextended. If this idea is correct, then the consequence of losing branches for an axon (such as the cyan one in Fig. 19.11) is that it can now dedicate more resources to its remaining multiply innervated junctions shared with other axons.

These results raise the obvious question of what resources might be in limited supply in an axon that could affect synapse maintenance. One popular idea is that the synaptic activity of an axon terminal is a key determinant in its ultimate fate. Is it possible that large motor units have fewer synaptic resources (e.g., synaptic vesicles per synapse) than small motor units? This idea was tested by generating mice in which a subset of axons had diminished choline acetyltransferase (ChAT, the enzyme that

synthesizes the acetylcholine; Buffelli et al., 2003). The results showed that when axons containing normal levels of ChAT co-innervated junctions with axons that had subnormal amounts, the subnormal axons typically occupied smaller territories. This result is consistent with the idea that axons that have the greatest number of synaptic branches may be at a disadvantage because they cannot maintain sufficient resources to drive postsynaptic cells as well as axons with fewer branches.

Summary

Structural studies of synapse elimination at the neuromuscular junction have provided a number of clues as to how and why branches are pruned. One important insight is that synapse elimination at individual neuromuscular junctions may be part of a larger-scale circuit optimization. This optimization may be occurring to ensure that all the connections that are maintained into adulthood are sufficiently strong that they can consistently drive the postsynaptic cell to threshold. Perhaps such an optimization requires axons to forfeit some of their synaptic connections to ensure that the ones that remain have sufficient resources to be consistently efficacious.

A ROLE FOR INTERAXONAL COMPETITION AND ACTIVITY

A number of lines of evidence suggest that the loss of synapses is the consequence of competition. The word *competition*, however, has many different meanings, and which of these definitions is most relevant to synaptic development is an area of some debate (Lichtman & Colman, 2000; Ribchester, 1992; Ribchester & Barry, 1994). In broad terms, competition occurs when more than one individual (in this case, more than one axon) is capable of having the same fate (e.g., sole occupation of a neuromuscular junction), and the probability of having that fate is related inversely to the number of such individuals. The point of defining competition so broadly is to emphasize that competition does not imply what kind of mechanism drives the outcome; even lotteries, where winners are picked by random, are competitions. Synapse elimination occurring on muscle fiber cells or cerebellar Purkinje cells is thought to be competitive because there is always only one axon remaining at the completion of the process, and therefore more than one axon cannot share the same fate.

Importantly, which particular axon is maintained does not seem to be preordained. There is no evidence of an extensive molecular specificity that uniquely matches each motor axon to an exclusive subset of muscle fibers or each climbing fiber to a matching set of Purkinje cells. Moreover, if there were some molecular blueprint, one might expect some stereotypy in that pattern between the same muscle in different individuals. However, painstaking axon tracing of all the motor units in a single muscle has revealed a complete lack of similarity in the organization of axons in many instantiations of the identical muscle (including in the left and right side of the same animal), arguing that no molecular patterning is establishing the particular wiring diagram. But if it is a competition that explains the variable outcomes, how is this competition being driven? Competitive mechanisms range from situations where the contestants interact with each other directly (e.g., a sumo wrestling match) to mechanisms where the contestants have little or nothing to do with each other, and a third party (a judge) decides the outcome (e.g., competition for a Pulitzer Prize). Evidence in the neuromuscular system suggests that synaptic competition may be more akin to the latter alternative, with the muscle fiber playing the role of judge.

Neuromuscular System

In 1970, Paul Redfern (1970) published the first physiological report of synapse elimination. He found that rat diaphragm muscle fibers were innervated by several motor neurons in the first postnatal week, but this multiple innervation was short lived; by two weeks of age, all muscle fibers were singly innervated (Fig. 19.1). At the time, he suggested that the extra innervation may be explained by the presence of motor neurons that project to a muscle in early life but subsequently undergo *cell death*. A few years later, this idea was shown to be incorrect when the tension elicited by activating single motor axons was shown to drop precipitously over the same period of early postnatal life, indicating that synapse elimination was accompanied by motor unit branch trimming rather than wholesale motor neuron loss (an event that occurs in the prenatal period) (Brown et al., 1976). Brown and colleagues also made the important point that because all neuromuscular junctions ended up with exactly one innervating axon, the elimination process was likely explained by interaxonal competition—otherwise all axons should leave a neuromuscular junction, or more than one converging axon should persist into adulthood. These two seminal reports stimulated many investigators to seek an understanding of the underlying mechanisms. Much of this effort has focused on the role of activity in synapse elimination. Many lines of evidence show that modifying neuromuscular activity has a large effect on synapse elimination (Thompson, 1985). However, it has remained unclear how activity might mediate interaxonal competition, as conflicting evidence suggested that either active (Ridge & Betz, 1984) or, surprisingly, inactive axons (Callaway, Soha, & Van Essen, 1987) are

at an advantage. At first glance, it may seem improbable that inactive axons could outcompete active ones (a result that is opposite to one discussed later, on the effects of monocular deprivation in visual system synapse elimination). Callaway et al. (1987) argued that the way muscles are used, the axons that are recruited most infrequently were the ones that in adults had the largest motor units. If their large sizes resulted from less branch trimming during development, then inactive axons might indeed have the advantage in synaptic competitions. Further complicating matters is the fact that the total number of axons innervating the muscle does not appear to change (Brown et al., 1976). This constancy implies that no individual axon's activity pattern could be considered *the* worst, because all axons maintain some neuromuscular junctions and thus all axons must outcompete other axons at some junctions.

Other experiments have questioned whether activity is even necessary for synapse elimination. For example, after reinnervation of adult muscle, evidence has been obtained showing that electrically silent inputs can displace other electrically silent inputs (Costanzo, Barry, & Ribchester, 2000). Moreover, in some cases, the presence of activity does not invariably lead to neuromuscular synapse elimination (Costanzo, Barry, & Ribchester, 1999).

As these conflicting experimental results demonstrate, the role of activity in synapse elimination is not easily summarized. One way to obtain a clearer idea of the role of activity is to view "activity" not as one phenomenon but as several different influences—for example:

- Synaptic competition may be affected by the relative efficacy (i.e., amplitude of the postsynaptic potential) of different axons attempting to drive the postsynaptic cell to threshold, with the most powerful input being favored.
- Alternatively, the firing frequency of an axon (i.e., action potentials per second) may have a negative impact, especially if the axon has a large arbor and thus insufficient resources to maintain efficacious synaptic transmission throughout its terminal branches.
- Finally, activity may affect the electrical properties of the postsynaptic cell or cell region, either encouraging synapse elimination (via postsynaptic activity) or preventing it by encouraging synapse formation (via the absence of postsynaptic activity) (Fig. 19.12).

Perhaps the most attractive hypothesis is that difference in activity between axons is critical for synapse elimination to occur. Indeed, because an axon has many synaptic release sites impinging on one postsynaptic cell, which do not seem to compete with themselves, interaxonal activity differences may be a crucial element of the competition. An experiment in which one part of a neuromuscular junction was desynchronized from the rest by silencing its acetylcholine receptors with alpha bungarotoxin (Box 19.1) examined how activity differences might give rise to synapse elimination. Focal postsynaptic silencing at one site within an otherwise normally active adult neuromuscular junction induced synapse withdrawal from the silenced site (Fig. 19.13). Postsynaptic silencing of an entire neuromuscular junction, however, did not cause synapse loss. These results suggested that synaptic activity at some synaptic sites can induce synaptic destabilization of other sites if they are not active at the same time. This view of the way activity modifies synaptic connections is related to the well-known theory of plasticity: *Hebb's postulate* (Hebb, 1949; see Chapter 52). Although Hebb argued that inputs that are consistently active when the postsynaptic cell is active are strengthened, this idea is the logical obverse: synapses that consistently fail to excite the postsynaptic cell when the cell is being activated are eliminated (Lichtman & Balice-Gordon, 1990; Stent, 1973). How might active synaptic sites destabilize silent ones? One suggestion is that active synapses generate two kinds of postsynaptic signals: one that protects them from the destabilizing effects of activity and one that punishes other inputs that are not active at the same time (Fig. 19.14). The physical basis of these protective and punishment signals remains unclear. Since neurotransmitter receptors are sometimes permeable to calcium, and the activity-induced depolarization can also raise intracellular calcium levels, one idea is that calcium signaling serves one or both of these roles.

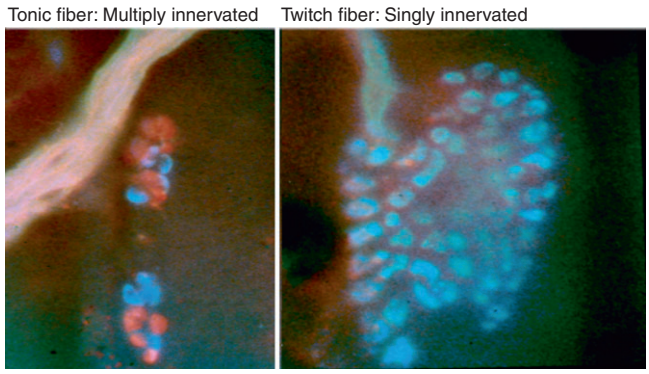


FIGURE 19.12 **Postsynaptic cells that don't fire action potentials don't undergo synapse elimination.** Photomicrographs of multiply and singly innervated snake muscle fibers. The singly innervated neuromuscular junctions are on twitch muscle fibers that have voltage-sensitive sodium channels. The multiply innervated neuromuscular junctions are found on tonic muscle fibers that do not have regenerative potentials. Labeling of different axons with different colors was accomplished by activity-dependent fluorescent labeling of axon terminals. *From Lichtman, Wilkinson, and Rich (1985).*

This view of the role of activity implies that if all axons were firing synchronously, then synapse elimination would not occur, which is exactly the conclusion reached using a regeneration model for synapse elimination (Busetto, Buffelli, Tognana, Bellico, & Cangiano, 2000). Interestingly, during the period of naturally occurring synapse elimination, there appears to be a switch from

BOX 19.1

α-BUNGAROTOXIN

A number of different dyes, stains, and markers are useful in revealing synaptic structure and function. Some of the more powerful have been borrowed from nature. Many toxins and poisons bind to specific proteins. One example is the snake toxin a-bungarotoxin (α-btx). This toxin is a constituent of the venom of a *Krait* snake of the species *bungarus*. The lethality of α-btx is the consequence of its ability to bind to the α subunits of nicotinic AChRs in skeletal muscle cells of vertebrates (with a few notable exceptions: snakes, mongooses, and hedgehogs). Because α-btx is an irreversible competitive antagonist of the AChR, its binding thus paralyzes and suffocates the prey. Researchers have taken advantage of this snake toxin to study many aspects of the AChR. For example, α-btx was used to purify AChRs from *Torpedo* membranes. Furthermore, the toxin can be conjugated to radioactive markers or to fluorescent molecules that can be seen in the microscope to stain receptors on muscle cells to determine their distribution, stability, and motility in the membrane. Because the toxin binds essentially irreversibly to the receptor in the muscle fiber membrane, receptors can be labeled once and then their behavior followed over time. This approach has provided substantial evidence about the stability of synaptic regions on muscle fibers, the lifetime of receptors in the membrane at the junctional sites, and how AChRs move within the plane of the membrane. For studies of synapse elimination in particular, the toxin also has been used to selectively inactivate some regions of a synapse by "puffing" it locally over a small region of a junction.

synchronous to asynchronous activity patterns (Buffelli, Busetto, Cangiano, & Cangiano, 2002; Personius & Balice-Gordon, 2001). It has been suggested that a gradual loss of electrical coupling among motor neurons may be the reason synapse elimination begins. This hypothesis was tested in mice that lack a gap junction protein (connexin 40) in which motor neuronal electrical coupling is reduced. In $Cx40^{-/-}$ muscles, synapse elimination was significantly accelerated, suggesting that asynchronous firing of neurons enhances synapse elimination (Personius, Chang, Mentis, O'Donovan, & Balice-Gordon, 2007).

Such an activity-based mechanism for synapse elimination would tend to pit the synchronously active synaptic terminals of one axon on a given postsynaptic cell (a *synaptic "cartel"*) against the terminals of other axons contacting the same target cell. In this way an axon's terminals on one postsynaptic cell cannot compete against themselves but rather serve as a competitive unit vying against the cartels of other axons.

Visual Cortex

Classic studies on the visual system by Hubel and Wiesel were the first to suggest that competition in fact was driving synaptic reorganization in the developing brain (Hubel & Wiesel, 1963; Hubel, Wiesel, & LeVay, 1977). In most species of young mammals, input neurons to layer IV in the visual cortex can be activated by inputs driven from both the left eye and the right eye—that is, they are driven binocularly. Subsequently, however, in many species, cortical input neurons become strongly dominated by either the right or left eye, but not both (Fig. 19.15). In agreement with this physiological result, the terminal arbors of the geniculocortical axons from the two eyes overlap in early postnatal life (Hubel et al., 1977; but see Horton & Hocking, 1996). However, with time, their projections appear to resolve into a striking pattern of alternating stripes known as *ocular dominance columns* (Fig. 19.15). This anatomical result was based on anterograde transneuronal transport of radioactive amino acid that was injected into one eye and passed through the thalamus to label the eye-specific axons in the optic radiation (Hubel et al., 1977). Despite the remarkable clarity of these stripes, their functional significance is not well understood, and in some mammals, such as rodents and new world monkeys, ocular dominance columns are absent (Horton & Adams, 2005; Livingstone, 1996).

During development, the ocular dominance columns' organization is less obvious because there is overlap in the thalamic inputs driven by the right and left eye to layer IV. Anatomically as development proceeds, each eye's columns become narrower and eventually almost nonoverlapping (especially in primates). Even though the narrowing in the widths of right and left eye ocular dominance columns is likely due to the loss of axonal branches in overlapping regions (Antonini & Stryker, 1993), this retraction of branches should not be taken to mean that the total number of thalamo-cortical synapses in layer IV is decreasing during this period. Similar to the peripheral nervous system, the axons associated with each eye elaborate many new synapses that more than compensate numerically for the lost connections of the withdrawing axons (Crowley & Katz, 2000; Erisir & Dreusicke, 2005). In other words, the process of ocular dominance column formation is one in which individual arbors lose synaptic connections with some targets but gain connections with others. Thus elimination restricts the neuronal population that is directly driven by each eye, but the synaptic

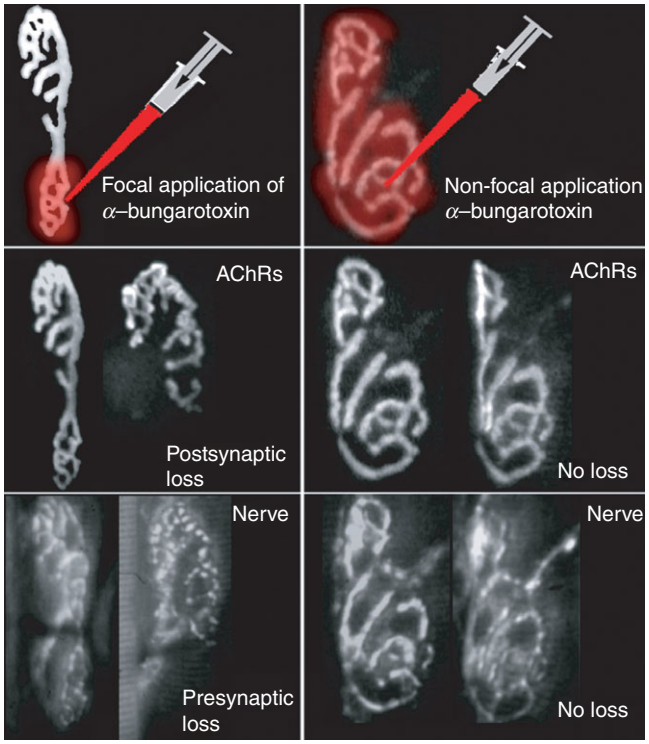


FIGURE 19.13 **An experimental test of the role of postsynaptic activity in synapse elimination.** If axons converging on the same postsynaptic cell were competing based on their efficacy in activating the postsynaptic cell, then locally blocking synaptic transmission at one axon's synaptic site should lead to its elimination and the maintenance of the other unblocked inputs. In this test, focal blockade of neuromuscular transmission was accomplished by applying saturating doses of α-bungarotoxin with a fine glass pipette (upper panels, red) to a neuromuscular junction in a living mouse. The AChR sites (middle panels) and the nerve terminals (lower panels) were viewed at the time of blockade and several times over the next few weeks. Following focal blockade (left panels), progressive loss of the nerve terminal staining and AChRs (arrows) was observed in regions previously saturated with α-bungarotoxin. However, when the entire neuromuscular junction was blocked (right panels), there was no loss. These results argue that active postsynaptic synaptic sites can cause the disassembly of inactive sites. *Adapted from Balice-Gordon and Lichtman (1994).*

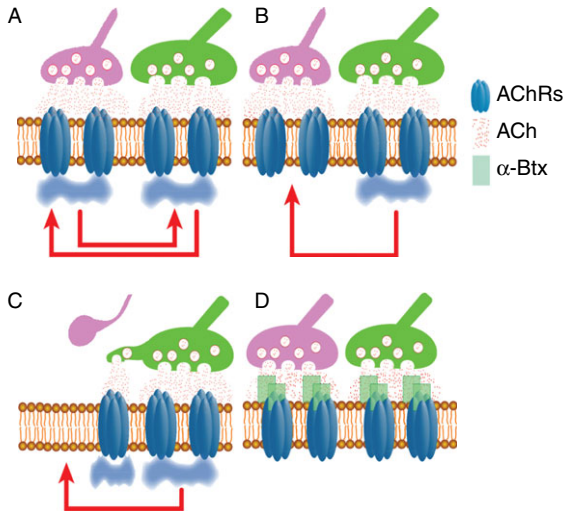


FIGURE 19.14 **Putative mechanism for a postsynaptic role in synapse elimination.** Postsynaptic receptor activation may elicit two opposing signals within a muscle fiber. One consequence of receptor activation is a "punishment" signal (here designated as red arrows) that causes destabilization of synaptic sites. Receptor activation may also generate a "protective" signal (here designated as blue clouds) that locally prohibits the punishment signals from destabilizing synapses in the vicinity of where receptor activation recently occurred. (A) When all the receptors are activated synchronously, as might occur before birth when motor neurons are electrically coupled (Personius & Balice-Gordon, 2001), there is no synaptic destabilization (due to protective blue clouds everywhere). (B) However, later in development, when two inputs are activated asynchronously, the active synaptic sites at any time point are protected (i.e., the synaptic sites beneath the active green axon are protected by a local blue cloud), whereas the asynchronously activated synapses (i.e., the synaptic sites under the pink axon) are not protected. (C) This asynchrony allows the more powerful input to destabilize the weaker one. This destabilization can lead to nerve and postsynaptic disappearance and/or nerve withdrawal, followed quickly by takeover of its former synaptic sites by the remaining axon, which would restabilize the site. (D) If all synaptic sites are inactive, as might occur with α-bungarotoxin application that inactivates nicotinic AChRs, then there are no blue clouds to protect synaptic sites, but also no red arrows to destabilize them. Thus synaptically silent and synchronously active synaptic sites (see panel A) do not undergo synapse elimination. *Idea adapted from Jennings (1994).*

addition strengthens the influence of one eye on the regions of cortex it continues to drive.

The most interesting aspect of this segregation is that the gradual removal of overlap in the two eyes' input streams leading to equally sized ocular dominance columns is not inevitable. Hubel and Wiesel showed that during a developmental *critical period* in early postnatal life, the widths of these columns can be dramatically and permanently changed by alterations in the relative amounts of visual experience in the two eyes. In particular, the outcome of the segregation can be radically skewed in favor of one eye if the activity of the other eye is decreased (e.g., by patching one eye). This monocular deprivation results in larger columns for the open eye and smaller columns for the deprived eye (Fig. 19.15). Once the critical period of sensitivity is passed (approximately the seventh postnatal week in kittens, the tenth week in ferrets, and the twentieth week in rhesus monkeys), the widths of the ocular dominance columns are fixed and no longer subject to shifts based on visual experience. Remarkably, even long-term monocular deprivation (of decades or more) apparently has little effect on the width of ocular dominance columns if the visual deprivation is begun after the critical period is over (approximately six years in humans; Keech & Kutschke, 1995). For example, in human patients in which one eye was removed for surgical reasons in adulthood or late childhood, postmortem analysis of the visual cortex indicated that eye removal has little or no effect on the width of the ocular dominance columns dominated by

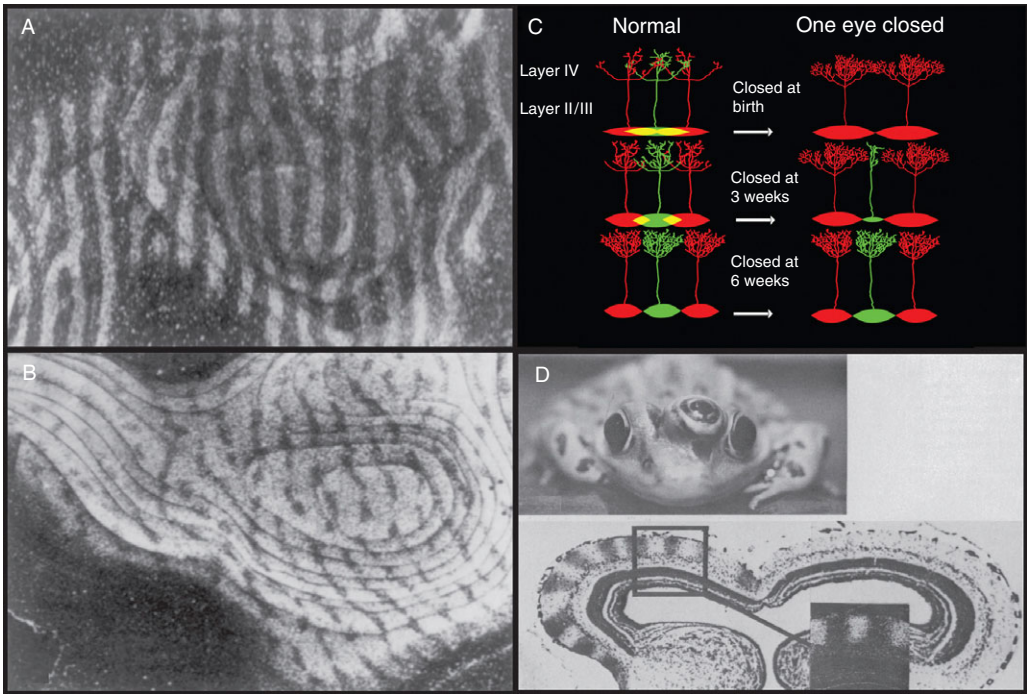


FIGURE 19.15 **Synapse elimination in the visual system.** (A) Ocular dominance columns of the neonatal monkey primary visual cortex in layer IVC, revealed by injecting [^{3}H]-proline into the vitreous of one eye. Light stripes (columns) represent sites containing the anterograde transported ^{3}H-amino acid from the injected eye. Dark regions are occupied by axons driven by the other eye. (B) Monocular deprivation by lid suture of one eye (2 weeks after birth for a period of 18 months) resulted in the shrinkage of the columns representing the deprived eye (dark stripes) and an expansion of the columns of the nondeprived eye (light stripes). (C) A schematic representation of ocular dominance column development represents the way in which a gradual segregation ocular dominance column could lead to the end of the critical period and progressively more modest effects of monocular deprivation as development ensues. (Top) At birth the afferents from the two eyes (red and green ovals) overlap completely in layer IV, and thus each eye is capable of maintaining inputs everywhere. At this young age, monocular deprivation would allow the nondeprived eye to remain in all parts of layer IV so that the entire cortex would be dominated by the red inputs. (Middle) In nondeprived animals, the two sets of afferents become progressively more segregated with age, meaning that by three weeks there would be regions of layer IV that are exclusively driven by the red or, as shown, green afferents. Once an eye's inputs are removed from a territory, it can no longer reoccupy that territory when the other eye is silenced. Hence monocular deprivation (that begins at three weeks) will spare a small strip of the inactive eye's territory (in this case the green regions). (Bottom) Once segregation is complete then monocular deprivation has no effect and the critical period is over. (D) A remarkable example showing how interactions between two eyes can cause segregation was found in frogs in which a third eye is implanted (at the tadpole stage) next to one of the eyes and projects with the native eye to the same optic tectum (ordinarily each frog tectum is monocular). After injection of H^3-proline into the normal eye, one optic tectum of a three-eyed frog shows dark and light bands strikingly similar to the ocular dominance columns observed in monkeys (D). *A–C adapted from Hubel et al. (1977); D from Constantine-Paton and Law (1978).*

the removed eye (Horton & Hocking, 1998). What might account for this dramatic change in sensitivity?

One idea is that during the critical period, thalamic afferents driven by the two eyes *compete* for control of the cortical neurons that they share temporarily. If each eye has the same average amount of activity, each ended up with similar amounts of cortical territory. However, if there were imbalances between the eyes in terms of visual experience, the outcome tipped the segregation in favor of one eye over the other. The skewing that resulted from depriving one eye of vision was due both to additional losses in the connections driven by the inactive eye (shrinking its ocular dominance columns) and to additional maintenance of the connections from the normally active nondeprived eye (maintaining its columns at the wider width it had at an earlier age) (Fig. 19.15). Ordinarily, each eye's afferents would relinquish its connections with approximately half of its postsynaptic target cells in visual cortex. Furthermore, binocular eye closure during the critical period appears to have far less serious effects than monocular occlusion. These results support the idea that synapse elimination is due to an activity-mediated competitive interaction between the connections driven by the two eyes. Because binocular deprivation has less dramatic effects on ocular dominance columns than monocular deprivation, here, as at the neuromuscular junction, active synaptic inputs seem to play a role in destabilizing inactive inputs.

Though these conclusions appear straightforward, the roles of activity in cortical refinements may need reevaluation as new investigations show that the mechanisms may be more complicated than originally imagined. In studies

of the rodent visual system, evidence suggests roles for both activity-dependent and -independent factors in developmental axonal refinements. Although mice lack ocular dominance columns, they do have a binocular cortical region that becomes progressively smaller as development proceeds. This shrinkage can be shifted with monocular deprivation during a critical period (Antonini, Fagiolini, & Stryker, 1999). This system has been used to show an important role of inhibitory circuits in both the establishment and maintenance of the critical period (Hensch, 2005). For example, deletion of the gene for glutamic acid decarboxylase, the enzyme that is responsible for the synthesis of the inhibitory neurotransmitter, GABA, prevents visual cortical refinements (Fagiolini & Hensch, 2000). Interestingly, these refinements can be reinitiated at any age by injecting GABA receptor agonists such as benzodiazepines into visual cortex. Thus intracortical inhibitory circuitry may be sufficient to trigger the opening or closure of the critical period in mice (Hensch, 2004; Hensch et al., 1998).

In mouse visual cortex, it has also been possible to study the sharpening of the retinotopic map in early postnatal life. The small receptive fields seen in adult visual cortex emerge from shrinkage of receptive field size in development (Issa, Trachtenberg, Chapman, Zahs, & Stryker, 1999). Activity plays a complex role in this sharpening; the effects of deprivation of formed vision (by lid suture) are different than the effects of pharmacological blockade or enucleation (Smith & Trachtenberg, 2007). A contralateral eye that is sensing light through a sutured lid impedes the refinement of the open eye's central projection, whereas either an open, or completely silent, contralateral eye does not. This result suggests that when inputs are synchronous (i.e., both eyes open), the activity-mediated refinements that shrink receptive fields occur more efficiently than when the same cortical neurons are receiving inputs with different activity patterns (i.e., in an animal with one open and one sutured eye). On the other hand, when one eye is entirely silent (i.e., by enucleation or pharmacological blockade), then the refinement of the open eye's inputs can still occur because in this case, there is no competing activity pattern from another eye. Thus some kinds of refinements may be mediated by cooperative interactions between the two eyes rather than competitive ones. One recent trend is that experiments that might have previously been interpreted strictly in terms of activity-mediated competition between different axons (à la Hubel and Wiesel) are now recast in terms of *homosynaptic* mechanisms of potentiation and depression or mechanisms of *synaptic homeostasis* (see Chapter 52). These newer frameworks for thinking about critical periods suggest that many different regulatory mechanisms working simultaneously may help to ensure that the right numbers and kinds of synapses survive the period of developmental refinements.

Thalamus

Separation of the inputs from the two eyes occurs twice in the visual system. Prior to the emergence of cortical ocular dominance columns, eye input to the lateral geniculate nucleus of the thalamus segregates into layers rather than columns. In embryonic cats, axon terminals of ganglion cells from the two eyes overlap extensively within the lateral geniculate nucleus before gradually segregating to form the characteristic *eye-specific* layers by birth. As in the cortex, this refinement process involves both the retraction of axonal branches from inappropriate regions of the geniculate nucleus and the elaboration of processes within the correct eye layer (Shatz, 1990). Physiological studies support anatomical observations that geniculate neurons initially are driven binocularly but maintain the axonal input from only one eye at maturity (Shatz, 1990). There is also a dramatic change in the convergence of retinal ganglion cell input to thalamic neurons related to a shrinkage in receptive field size (Chen & Regehr, 2000; Tavazoie & Reid, 2000) (Fig. 19.16).

It is likely that *spontaneous activity* as opposed to actual visual experience is important in the segregation of retinogeniculate connections in the thalamus. This view is based on the fact that in cats, ferrets, and monkeys, the eye-specific layers are established well before the retina is sensitive to light and is dependent on the spontaneous activity

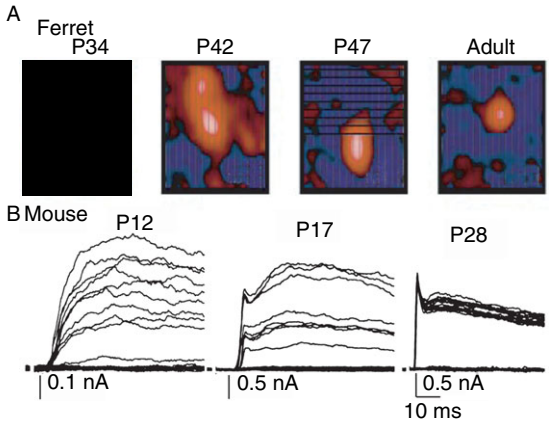


FIGURE 19.16 **Synapse elimination in the lateral geniculate nucleus of the thalamus.** (A) In the ferret lateral geniculate nucleus, receptive fields of geniculate neurons are larger and much more diffuse at one month of age than those receptive fields observed in adult neurons. Red regions represent the receptive field map of geniculate neurons that correspond to areas excited by bright stimuli. Note that red areas become smaller as development proceeds. The shrinkage in the receptive field is likely to result from the elimination of the convergence of multiple retinal afferents onto each geniculate neuron (B). At P12 in mouse, multiple retinogeniculate axons are recruited as stimulation intensities to the bundle of axons is increased (see also Fig. 19.1C). At P17 there are fewer steps, and after P28, only one or two inputs innervate each geniculate neuron (i.e., no steps in the evoked-synaptic currents are seen even though optic nerve stimulation is increased). *(A) Adapted from Tavazoie and Reid (2000). (B) Adapted from Chen and Regehr (2000).*

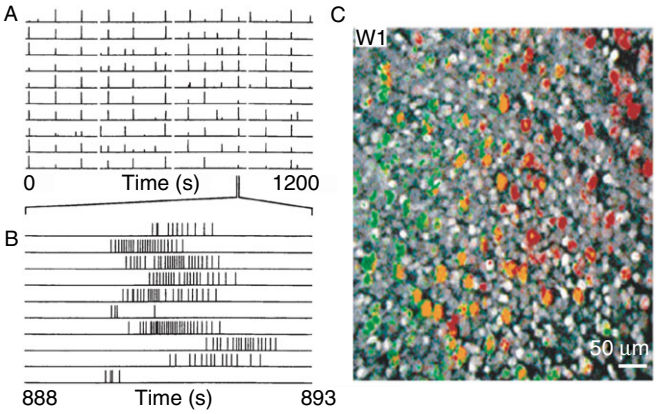


FIGURE 19.17 **Immature retinal ganglion cells show correlated patterns of activity.** (A) Using an array of extracellular recordings, rhythmic bursts of action potentials (indicated by vertical lines) are synchronized between neighboring retinal ganglion cells before eye opening in ferret. (B) Action potential bursts expanded in time scale show that each burst corresponds to ten or more action potentials. (C) Using calcium indicators it has been possible to observe waves of neuronal activity in immature retinal ganglion cells (pseudo-colored image). In this example, a wave of neuronal activity propagates through the retina. Pseudo-colored cells indicating the temporal firing pattern of retinal ganglion neurons; cells in green fire before yellow ones, and lastly red cells. *(A, B) from Meister, Wong, Baylor, and Shatz (1991). (C) from Wong (1999).*

of these inputs. What information may be contained in the spontaneous activity patterns of immature retinas that could lead to eye-specific lamination? Electrophysiological recordings and Ca^{++} imaging studies demonstrate that each immature retina generates correlated propagating *waves* (Fig. 19.17), which have no preferred direction of propagation and that occur periodically, about once a minute. It is possible that retinal waves contain temporal and spatial cues that guide activity-dependent refinement of retinogeniculate connections. For example, because waves are generated independently in each retina, activities from the two eyes are unlikely to be coincident, while nearby retinal ganglion cells in the same eye would tend to have similar spontaneous activity patterns. Asynchrony between the inputs of the two eyes could account for the segregation of inputs into different eye-specific layers in the thalamus. Since ocular dominance column formation is initiated before birth, the spontaneous patterns of activity from the two eyes could be responsible for the eye-specific pathways throughout the visual system. Moreover, because the waves ensure that nearby retinal ganglion cells are better synchronized than more distant cells, geniculate neurons are able to gauge neighbor relationships in the retina by their sequential activation. This feature could be useful for refinements of the retinotopic map in both thalamus and cortex. Pharmacological blockade of retinal waves during the period of eye-specific segregation prevents the emergence of these layers, suggesting that activity from the retinas is involved (Wong, 1999). However, these activity patterns must be only part of the story: the laminar organization of the lateral geniculate is stereotyped from animal to animal, suggesting that other developmental mechanisms are also at play.

Cerebellum

A particularly clear example of synapse elimination in the developing CNS is the *climbing fiber input* onto cerebellar Purkinje cells. Climbing fibers are the terminals of the axons arriving from inferior olive neurons that form strong synaptic connections to Purkinje cells. In adults only one climbing fiber innervates each Purkinje cell. That input might contain 500 synaptic boutons that tightly invest the large ascending proximal dendrite and its major branches. Immature climbing fibers, on the other hand, form fewer synapses, mostly on the Purkinje cell soma, and many climbing fibers project to each Purkinje cell in the first postnatal week (in rodents). The transition from multiple innervation to single innervation of individual Purkinje cells occurs at the same time there is a change in the number of Purkinje cells innervated by each climbing fiber. This "neural unit" shrinkage is seemingly quite analogous to the reduction in the size of motor units in the peripheral nervous system. Thus one olivocerebellar axon may give rise to branches that innervate more than 100 Purkinje cells during the first postnatal week, but over the next several weeks its projection is trimmed to only about 7 Purkinje cells (Fig. 19.18), albeit these connections are much more powerful.

It has long been appreciated that the loss of climbing fiber inputs depends on the presence of parallel fiber innervation from granule cells of the distal part of the Purkinje cell arbor. Elimination of granule cell inputs to Purkinje cells by X irradiation, viral infection, or in mutants such as *reeler*, *weaver*, and *staggerer* results in a higher incidence of Purkinje cells that are multiply innervated by climbing fibers in adulthood. Some studies suggest that the activity of the parallel fiber input is the important parameter. Perturbation of activity along the parallel fiber–Purkinje cell pathway in mGluR1 and GluRδ2 knockout mice or application of NMDA receptor antagonists all inhibit the elimination of climbing fibers (Hashimoto & Kano, 2005). In addition, disruption of one calcium binding kinase (PKC gamma) appears to selectively prevent climbing fiber elimination (Hashimoto & Kano, 2005). Although the mechanism by which this kinase alters synapse elimination is not known, these animals recently have been shown to have a profound deficit in vestibulo-ocular reflex (VOR) motor learning but not other kinds of cerebellar learning (Kimpo & Raymond, 2007). These results imply that synapse elimination may be important in generating the circuitry for some kinds of adult learning. As we will mention later, however, synapse elimination itself may be a form of learning.

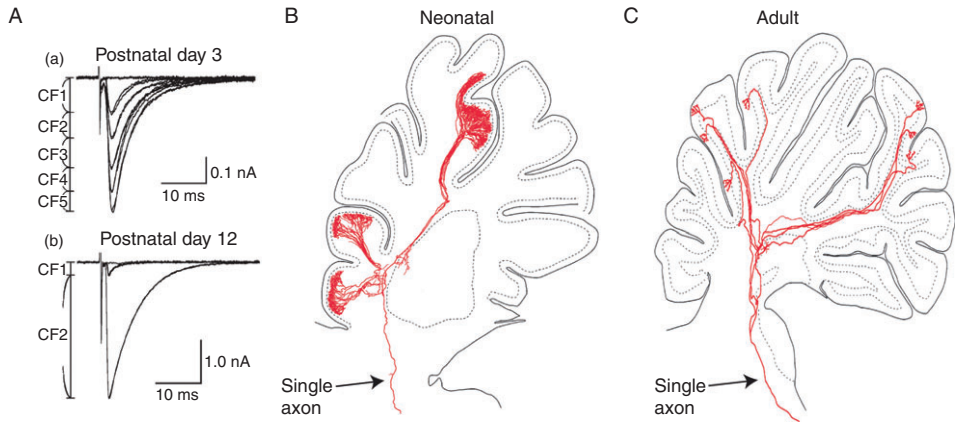


FIGURE 19.18 **Synapse elimination and axonal pruning of climbing fibers in the neonatal period.** (A) Recordings from Purkinje cells while olivocerebellar axons are stimulated in the inferior olive show functional evidence of synapse elimination. At birth, each Purkinje cell is innervated by several different climbing fibers, and as the strength of stimulation is increased, additional inputs are recruited (compare with Figs. 19.1C and 19.15B). The average number of climbing fibers innervating each cerebellar Purkinje cell, in the rat, decreases gradually as the animal matures until most are singly innervated. (B, C) Reconstructions of the trajectory of single neonatal (B) and adult (C) olivocerebellar climbing fiber axons. Both neonatal and adult axons terminate in several separate lobules in the hemisphere. However, neonatal olivocerebellar axons have many more branches than those in adult axons and presumably innervate many more Purkinje cells than those in adult animals. After pruning is complete, each axon gives rise to ~7 climbing fibers that each singly innervates a different Purkinje cell. *(A) Adapted from Hashimoto and Kano (2003). (B, C) Modified from Sugihara (2005).*

Summary

In many parts of the central and peripheral nervous system, the divergence and convergence of synaptic circuits are decreased. A popular hypothesis, albeit still somewhat perplexing, is that neural activity differences between axons underly synapse elimination. This hypothesis has driven researchers to try many kinds of experiments to test electrical activity's role. At present, the precise way in which electrical activity exerts an influence on the synapse elimination process remains a central and unresolved question.

IS SYNAPSE ELIMINATION STRICTLY A DEVELOPMENTAL PHENOMENON?

In this chapter, we presented evidence to demonstrate that synapse elimination is a powerful force that can refine synaptic circuits in young animals, based on interneuronal competition. Is there any reason to think it is more than a strictly developmental phenomenon? The most important form of adult plasticity must certainly be memory. Might synapse elimination have something to do with memory? A number of neurobiologists, including Kandel (1967), Toulouse, Dehaene, and Changeux (1986), and Edeleman (1988), have explicitly made arguments for selection (as opposed to instruction) as potentially playing an important role in learning. The idea is that in the brain synaptic circuitry exists *a priori* for many things that may ultimately be learned, so that learning might occur by the selection of synaptic pathways that already exist rather than construction of new circuits. Although such selection could occur by increasing the strength of one set of synaptic interconnections or weakening of others, it could also occur by completely eliminating some circuits. It is important to emphasize the distinction between plasticity that alters the strengths of existing connections and the more extreme kind of plasticity, analogous to the developmental synaptic processes described here, that causes permanent eradication of an axon's input to particular postsynaptic cells.

Because postsynaptic cells appear to be the intermediary in synaptic competition leading to axonal removal (see earlier), once an axon's synaptic drive to a postsynaptic cell is removed, it can no longer have any influence on the synaptic connections of the other axons that remain connected. Complete loss of influence following synaptic disconnection is thus a plausible explanation for the finite length of critical periods. For example, once all the inputs driven by an eye deprived of vision are eliminated, return of visual experience in that eye can no longer cause a shift in ocular dominance columns if that shift is mediated via activation of postsynaptic cells. The same argument could also be made for memory. Memories have a kind of indelibility that prevents more recent unrelated memories from "overwriting" prior ones. Input elimination is an attractive means of ensuring indelibility because by eliminating competing (i.e., asynchronously firing) inputs, a circuit becomes sheltered from disruption by different activity patterns.

A model of memory based on this kind of synapse elimination, however, would require that axonal inputs continue to be eliminated in the adult brain. That critical periods in the visual system are strictly developmental can be used as an argument against the idea that these kinds of changes may underlie adult memory. On the other hand, critical periods in the visual system tend to be prolonged in proportion to their distance from the retinal

input. For example, critical periods for higher visual processing areas occur later in development than in those areas that are more proximal in the visual pathway. The loss of overlapping connections is known to occur prenatally within thalamic circuitry well before segregation in layer IV primary visual cortex, and higher anatomical levels in the visual cortex that receive input from layer IV segregate out later than layer IV. A possible explanation for this sequential crystallization of brain regions may be that synapse elimination can occur only when a cohort of synchronous inputs work together to drive the elimination of competing inputs. Such a collection of synchronously active neurons requires that the presynaptic input to these cells has itself sorted out. It also is the case that the length of the critical periods for vision are vastly longer in humans than other mammals, as is the rest of our neotenic development. For example, whereas our closest animal relatives finish the critical period for monocular deprivation by 7 months of age, in humans, monocular deprivation can affect visual acuity even in children 6 to 7 years old.

Summary

We would not like to give the impression that naturally occurring synapse elimination at developing systems is the equivalent of learning and memory. But, as neurobiologists who have studied this phenomenon and mulled these ideas over for many years, we have come to the conclusion that permanent loss of axonal input is an attractive mechanism for information storage. Whether our bias is reasonable based on the data or rather due to structural elimination of competing hypotheses from our brains, we do not know.

References

Antonini, A., Fagiolini, M., & Stryker, M. P. (1999). Anatomical correlates of functional plasticity in mouse visual cortex. *Journal of Neuroscience, 19,* 4388–4406.

Antonini, A., & Stryker, M. P. (1993). Rapid remodeling of axonal arbors in the visual cortex. *Science, 260,* 1819–1821.

Balice-Gordon, R. J., Chua, C. K., Nelson, C. C., & Lichtman, J. W. (1993). Gradual loss of synaptic cartels precedes axon withdrawal at developing neuromuscular junctions. *Neuron, 11,* 801–815.

Balice-Gordon, R. J., & Lichtman, J. W. (1994). Long-term synapse loss induced by focal blockade of postsynaptic receptors. *Nature, 372,* 519–524.

Bingol, B., Wang, C. F., Arnott, D., Cheng, D., Peng, J., & Sheng, M. (2010). Autophosphorylated CaMKII alpha acts as a scaffold to recruit proteasomes to dendritic spines. *Cell, 140,* 567–578.

Bishop, D. L., Misgeld, T., Walsh, M. K., Gan, W. B., & Lichtman, J. W. (2004). Axon branch removal at developing synapses by axosome shedding. *Neuron, 44,* 651–661.

Boeke, J. (1932). *Nerve endings, motor and sensory* (Vol. 1). New York: Hafner Press.

Brown, M. C., Jansen, J. K., & Van Essen, D. (1976). Polyneuronal innervation of skeletal muscle in new-born rats and its elimination during maturation. *Journal of Physiology, 261,* 387–422.

Buffelli, M., Burgess, R. W., Feng, G., Lobe, C. G., Lichtman, J. W., & Sanes, J. R. (2003). Genetic evidence that relative synaptic efficacy biases the outcome of synaptic competition. *Nature, 424,* 430–434.

Buffelli, M., Busetto, G., Cangiano, L., & Cangiano, A. (2002). Perinatal switch from synchronous to asynchronous activity of motoneurons: link with synapse elimination. *Proceedings of National Academy of Sciences of the United States of America, 99,* 13200–13205.

Busetto, G., Buffelli, M., Tognana, E., Bellico, F., & Cangiano, A. (2000). Hebbian mechanisms revealed by electrical stimulation at developing rat neuromuscular junctions. *Journal of Neuroscience, 20,* 685–695.

Callaway, E. M., Soha, J. M., & Van Essen, D. C. (1987). Competition favouring inactive over active motor neurons during synapse elimination. *Nature, 328,* 422–426.

Chen, C., & Regehr, W. G. (2000). Developmental remodeling of the retinogeniculate synapse. *Neuron, 28,* 955–966.

Colman, H., Nabekura, J., & Lichtman, J. W. (1997). Alterations in synaptic strength preceding axon withdrawal. *Science, 275,* 356–361.

Constantine-Paton, M., & Law, M. I. (1978). Eye-specific termination bands in tecta of three-eyed frogs. *Science, 202,* 639–641.

Costanzo, E. M., Barry, J. A., & Ribchester, R. R. (1999). Co-regulation of synaptic efficacy at stable polyneuronally innervated neuromuscular junctions in reinnervated rat muscle. *Journal of Physiology, 521* (Pt 2), 365–374.

Costanzo, E. M., Barry, J. A., & Ribchester, R. R. (2000). Competition at silent synapses in reinnervated skeletal muscle. *Nature Neuroscience, 3,* 694–700.

Crowley, J. C., & Katz, L. C. (2000). Early development of ocular dominance columns. *Science, 290,* 1321–1324.

Ding, M., Chao, D., Wang, G., & Shen, K. (2007). Spatial regulation of an E3 ubiquitin ligase directs selective synapse elimination. *Science, 317,* 947–951.

Eaton, B. A., & Davis, G. W. (2003). Synapse disassembly. *Genes & Development, 17,* 2075–2082.

Edeleman, G. (1988). *The theory of neuronal group selection*. New York: Basic Books.

Erisir, A., & Dreusicke, M. (2005). Quantitative morphology and postsynaptic targets of thalamocortical axons in critical period and adult ferret visual cortex. *Journal of Comparative Neurology, 485,* 11–31.

Fagiolini, M., & Hensch, T. K. (2000). Inhibitory threshold for critical-period activation in primary visual cortex. *Nature, 404,* 183–186.

Feng, G., Mellor, R. H., Bernstein, M., Keller-Peck, C., Nguyen, Q. T. Wallace, M. et al. (2000). Imaging neuronal subsets in transgenic mice expressing multiple spectral variants of GFP. *Neuron, 28,* 41–51.

Gan, W. B., & Lichtman, J. W. (1998). Synaptic segregation at the developing neuromuscular junction. *Science, 282,* 1508–1511.

Gorio, A., Marini, P., & Zanoni, R. (1983). Muscle reinnervation—III. Motoneuron sprouting capacity, enhancement by exogenous gangliosides. *Neuroscience, 8,* 417–429.

Hashimoto, K., & Kano, M. (2003). Functional differentiation of multiple climbing fiber inputs during synapse elimination in the developing cerebellum. *Neuron, 38,* 785–796.

Hashimoto, K., & Kano, M. (2005). Postnatal development and synapse elimination of climbing fiber to Purkinje cell projection in the cerebellum. *Neuroscience Research, 53,* 221–228.

Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.

Hensch, T. K. (2004). Critical period regulation. *Annual Review of Neuroscience, 27,* 549–579.

Hensch, T. K. (2005). Critical period plasticity in local cortical circuits. *Nature Reviews Neuroscience, 6,* 877–888.

Hensch, T. K., Fagiolini, M., Mataga, N., Stryker, M. P., Baekkeskov, S., & Kash, S. F. (1998). Local GABA circuit control of experience-dependent plasticity in developing visual cortex. *Science, 282,* 1504–1508.

Horton, J. C., & Adams, D. L. (2005). The cortical column: A structure without a function. *Philosophical Transactions of Royal Society of London. Series B, Biological Sciences, 360,* 837–862.

Horton, J. C., & Hocking, D. R. (1996). An adult-like pattern of ocular dominance columns in striate cortex of newborn monkeys prior to visual experience. *Journal of Neuroscience, 16,* 1791–1807.

Horton, J. C., & Hocking, D. R. (1998). Effect of early monocular enucleation upon ocular dominance columns and cytochrome oxidase activity in monkey and human visual cortex. *Visual Neuroscience, 15,* 289–303.

Hubel, D. H., & Wiesel, T. N. (1963). Receptive fields of cells in striate cortex of very young, visually inexperienced kittens. *Journal of Neurophysiology, 26,* 994–1002.

Hubel, D. H., Wiesel, T. N., & LeVay, S. (1977). Plasticity of ocular dominance columns in monkey striate cortex. *Philosophical Transactions of Royal Society of London. Series B, Biological Sciences, 278,* 377–409.

Issa, N. P., Trachtenberg, J. T., Chapman, B., Zahs, K. R., & Stryker, M. P. (1999). The critical period for ocular dominance plasticity in the Ferret's visual cortex. *Journal of Neuroscience, 19,* 6965–6978.

Jackson, H., & Parks, T. N. (1982). Functional synapse elimination in the developing avian cochlear nucleus with simultaneous reduction in cochlear nerve axon branching. *Journal of Neuroscience, 2,* 1736–1743.

Jennings, C. (1994). Developmental neurobiology. Death of a synapse. *Nature, 372,* 498–499.

Kalinovsky, A., Boukhtouche, F., Blazeski, R., Bornmann, C., Suzuki, N., Mason, C. A., & Scheiffele, P. (2011). Development of axon-target specificity of ponto-cerebellar afferents. *PLoS Biology, 9,* e1001013.

Kandel, E. (1967). *Cellular studies of learning*. New York: Rockefeller University Press.

Kasthuri, N., & Lichtman, J. W. (2003). The role of neuronal identity in synaptic competition. *Nature, 424,* 426–430.

Keech, R. V., & Kutschke, P. J. (1995). Upper age limit for the development of amblyopia. *Journal of Pediatric Ophthalmology & Strabismus, 32,* 89–93.

Keller-Peck, C. R., Walsh, M. K., Gan, W. B., Feng, G., Sanes, J. R., & Lichtman, J. W. (2001). Asynchronous synapse elimination in neonatal motor units: studies using GFP transgenic mice. *Neuron, 31,* 381–394.

Kimpo, R. R., & Raymond, J. L. (2007). Impaired motor learning in the vestibulo-ocular reflex in mice with multiple climbing fiber input to cerebellar Purkinje cells. *Journal of Neuroscience, 27,* 5672–5682.

Laskowski, M. B., Colman, H., Nelson, C., & Lichtman, J. W. (1998). Synaptic competition during the reformation of a neuromuscular map. *Journal of Neuroscience, 18,* 7328–7335.

Li, Z., Jo, J., Jia, J. M., Lo, S. C., Whitcomb, D. J., Jiao, S. et al. (2010). Caspase-3 activation via mitochondria is required for long-term depression and AMPA receptor internalization. *Cell, 141,* 859–871.

Lichtman, J. W. (1977). The reorganization of synaptic connexions in the rat submandibular ganglion during post-natal development. *Journal of Physiology, 273,* 155–177.

Lichtman, J. W., & Balice-Gordon, R. J. (1990). Understanding synaptic competition in theory and in practice. *Journal of Neurobiology, 21,* 99–106.

Lichtman, J. W., & Colman, H. (2000). Synapse elimination and indelible memory. *Neuron, 25,* 269–278.

Lichtman, J. W., & Purves, D. (1980). The elimination of redundant preganglionic innervation to hamster sympathetic ganglion cells in early post-natal life. *Journal of Physiology, 301,* 213–228.

Lichtman, J. W., Wilkinson, R. S., & Rich, M. M. (1985). Multiple innervation of tonic endplates revealed by activity-dependent uptake of fluorescent probes. *Nature, 314,* 357–359.

Livingstone, M. S. (1996). *Ocular dominance columns in New World monkeys. Journal of Neuroscience, 16,* 2086–2096.

Lu, T., & Trussell, L. O. (2007). Development and elimination of end-bulb synapses in the chick cochlear nucleus. *Journal of Neuroscience, 27,* 808–817.

Mariani, J. (1983). Elimination of synapses during the development of the central nervous system. *Progress in Brain Research, 58,* 383–392.

Meister, M., Wong, R. O., Baylor, D. A., & Shatz, C. J. (1991). Synchronous bursts of action potentials in ganglion cells of the developing mammalian retina. *Science, 252,* 939–943.

Personius, K. E., & Balice-Gordon, R. J. (2001). Loss of correlated motor neuron activity during synaptic competition at developing neuromuscular synapses. *Neuron, 31,* 395–408.

Personius, K. E., Chang, Q., Mentis, G. Z., O'Donovan, M. J., & Balice-Gordon, R. J. (2007). Reduced gap junctional coupling leads to uncorrelated motor neuron firing and precocious neuromuscular synapse elimination. *Proceedings of National Academy of Sciences of the United States of America, 104,* 11808–11813.

Redfern, P. A. (1970). Neuromuscular transmission in new-born rats. *Journal of Physiology, 209,* 701–709.

Ribchester, R. R. (1992). Cartels, competition and activity-dependent synapse elimination. *Trends in Neurosciences, 15,* 389. author reply 390–381.

Ribchester, R. R., & Barry, J. A. (1994). Spatial versus consumptive competition at polyneuronally innervated neuromuscular junctions. *Experimental Physiology, 79,* 465–494.

Ridge, R. M., & Betz, W. J. (1984). The effect of selective, chronic stimulation on motor unit size in developing rat muscle. *Journal of Neuroscience, 4,* 2614–2620.

Shatz, C. J. (1990). Impulse activity and the patterning of connections during CNS development. *Neuron, 5,* 745–756.

Smith, S. L., & Trachtenberg, J. T. (2007). Experience-dependent binocular competition in the visual cortex begins at eye opening. *Nature Neuroscience, 10,* 370–375.

Stent, G. S. (1973). A physiological mechanism for Hebb's postulate of learning. *Proceedings of National Academy of Sciences of the United States of America, 70,* 997–1001.

Sugihara, I. (2005). Microzonal projection and climbing fiber remodeling in single olivocerebellar axons of newborn rats at postnatal days 4–7. *Journal of Comparative Neurology, 487,* 93–106.

Tavazoie, S. F., & Reid, R. C. (2000). Diverse receptive fields in the lateral geniculate nucleus during thalamocortical development. *Nature Neuroscience, 3,* 608–616.

Thompson, W. J. (1985). Activity and synapse elimination at the neuromuscular junction. *Cell and Molecular Neurobiology, 5,* 167–182.

Toulouse, G., Dehaene, S., & Changeux, J. P. (1986). Spin glass model of learning by selection. *Proceedings of National Academy of Sciences of the United States of America, 83,* 1695–1698.

Walsh, M. K., & Lichtman, J. W. (2003). In vivo time-lapse imaging of synaptic takeover associated with naturally occurring synapse elimination. *Neuron, 37,* 67–73.

Wang, G. Y., Liets, L. C., & Chalupa, L. M. (2001). Unique functional properties of on and off pathways in the developing mammalian retina. *Journal of Neuroscience, 21,* 4310–4317.

Wiesel, T. N. (1982). Postnatal development of the visual cortex and the influence of environment. *Nature, 299,* 583–591.

Wong, R. O. (1999). Retinal waves and visual system development. *Annual Review of Neuroscience, 22,* 29–47.

Juan C. Tapia and Jeff W. Lichtman

This page intentionally left blank

CHAPTER

20

Dendritic Development

Dendrites play a critical role in information processing in the nervous system as substrates for synapse formation and signal integration. Neurons have highly branched, cell type–specific dendritic arbors (or dendritic trees). The size, shape, branching pattern, and position of dendritic arbors are the main determinants of what inputs they receive and how those inputs are integrated. It is now widely recognized that dendrites develop in constant interaction with other neurons and glia. The signals from these other cells affect dendritic arbor development in different spatial domains and time scales. For instance, synaptic inputs may increase calcium influx rapidly and locally to enhance rates of branch addition and stabilization. This dynamic morphological remodeling allows neurons to modify their structure and synaptic connectivity in response to afferent inputs (Cline, 2001; Ruthazer, Akerman, & Cline, 2003; Sin, Haas, Ruthazer, & Cline, 2002). In contrast, calcium signals with slower temporal dynamics may selectively signal to the nucleus to trigger gene transcription (Dolmetsch, Pajvani, Fife, Spotts, & Greenberg, 2001; Kornhauser et al., 2002; Wu, Deisseroth, & Tsien, 2001). Such activity-induced genes can then have profound effects on dendritic arbor structure and function (Cantallops & Cline, 2008; Leslie & Nedivi, 2011; Nedivi, 1999; Nedivi, Wu, & Cline, 1998). Dendrite arbor structure and plasticity are altered under a variety of neurological disorders, including mental retardation (Bagni & Greenough, 2005; Govek et al., 2004; Na & Monteggia, 2011; Newey, Velamoor, Govek, & Van Aelst, 2005) or seizure (Nishimura, Gu, & Swann, 2011), and can be affected by exposure to drugs, including nicotine (Gonzalez, Gharbawie, Whishaw, & Kolb, 2005) and cocaine (Kolb, Gorny, Li, Samaha, & Robinson, 2003; Morrow, Elsworth, & Roth, 2005). The study of dendritic arbor development can therefore provide important insight into the cellular basis of normal brain development, as well as neurological and psychiatric disorders.

After a brief description of dendritic arbor development, we discuss four principles of dendritic development: (1) Transcriptional programs establish rudimentary dendritic morphology characteristic of distinct neuronal subclasses. (2) Extracellular cues interact with signaling pathways to sculpt dendritic morphology. (3) Coordinate sculpting of axonal and dendritic arbors leads to formation of neural circuits optimized for information processing. (4) Electrical and synaptic activity refine dendritic arbor morphology and synaptic connectivity.

Although these principles are general ones, applicable to most if not all parts of the nervous system, we illustrate each one with reference to a model system that has been useful for studying it. Thus, we highlight *Drosophila* mechanoreceptor neurons for genetic analysis, mammalian cortex for extrinsic signals, vertebrate retina for circuit formation, and *Xenopus retinotectal* system for activity-dependent structural plasticity.

DYNAMICS OF DENDRITIC ARBOR DEVELOPMENT

Dendritic arbor development requires global, arbor-wide architectural modifications (e.g., generalized arbor growth), as well as localized structural changes (e.g., sprouting and retraction of higher order branches). Some changes in dendritic arbor morphology occur rapidly in response to synaptic inputs or growth factors, whereas others occur with some time delay, secondary to new gene transcription and protein turnover.

The advent of *in vivo* time-lapse imaging has allowed the assessment of cellular events underlying the development of dendritic arbors. *In vivo* imaging of dendritic arbor development was pioneered in the 1980s by Purves and colleagues, who discovered that peripheral neurons could be imaged repeatedly in developing mice (Hume & Purves, 1981).

© 2013 Elsevier Inc. All rights reserved.

These studies revealed considerable heterogeneity in the dynamics of dendritic arbor structure and suggested that dendrite structure was affected by differences in afferent inputs to different neurons. More than a decade passed before the methods were developed to reliably label single neurons in the CNS and image their dendrites over time without damage (Wu & Cline, 1998). *In vivo* time-lapse imaging of neurons is now routine in several systems, including fish, amphibian tadpoles, *Drosophila*, and rodents, thanks to technical advances in imaging reagents and widespread use of 2-photon microscopy.

In vivo time-lapse imaging studies of developing optic tectal neurons in *Xenopus* and Zebrafish and peripheral sensory neurons in *Drosophila* demonstrate that the dendritic arbors develop through a gradual process in which branches are added and retracted rapidly (Fig. 20.1). Figure 20.1A illustrates the generalized growth of differentiating optic tectal neurons imaged in the intact *Xenopus* tadpole over three days, and Figure 20.1B shows an optic tectal neuron imaged *in vivo* over about one day, but at shorter intervals. Widespread changes in dendritic architecture are apparent when time-lapse images are collected over daily intervals or longer, while short interval imaging unveils sites of local branch dynamics and growth that contribute to the changes in structure seen at longer intervals. Many more branches are added to the arbor than are ultimately maintained, so that the net elaboration of the dendritic arbor occurs as a result of the stabilization of a tiny fraction of the newly added branches. These branches then become the substrate for further branch additions. In this way the complex arbor develops gradually as a result of concurrent and iterative branch addition, retraction, stabilization, and extension (Cline, 2001; Hua & Smith, 2004). Time-lapse imaging studies in other systems including rodent hippocampus and cortex, and chick auditory system (Dailey & Smith, 1996; Sorensen & Rubel, 2006) indicate that this plan is fundamental to dendrite development.

TRANSCRIPTIONAL CONTROL OF DENDRITE DEVELOPMENT

Studies using *Drosophila* genetics have been instrumental in the identification of core transcriptional programs that control dendrite development. The dendritic arborization (da) neurons, a group of *Drosophila* sensory neurons with a stereotyped dendritic branching pattern, have provided a useful assay system for the genetic dissection of dendrite development (Gao, Brenman, Jan, & Jan, 1999; Grueber, Jan, & Jan, 2002; Jan & Jan, 2010). Each hemisegment of the abdomen of the *Drosophila* embryo or larva has 15 da neurons that can be subdivided into four classes based on their stereotyped location and unique dendritic branching pattern (Fig. 20.2) (Grueber, Jan, & Jan, 2002; Grueber, Ye, Moore, Jan, & Jan, 2003b). Class I and II have relatively simple dendritic branching patterns and small dendritic fields, whereas class III and IV neurons have more complex dendritic branching patterns and large dendritic fields.

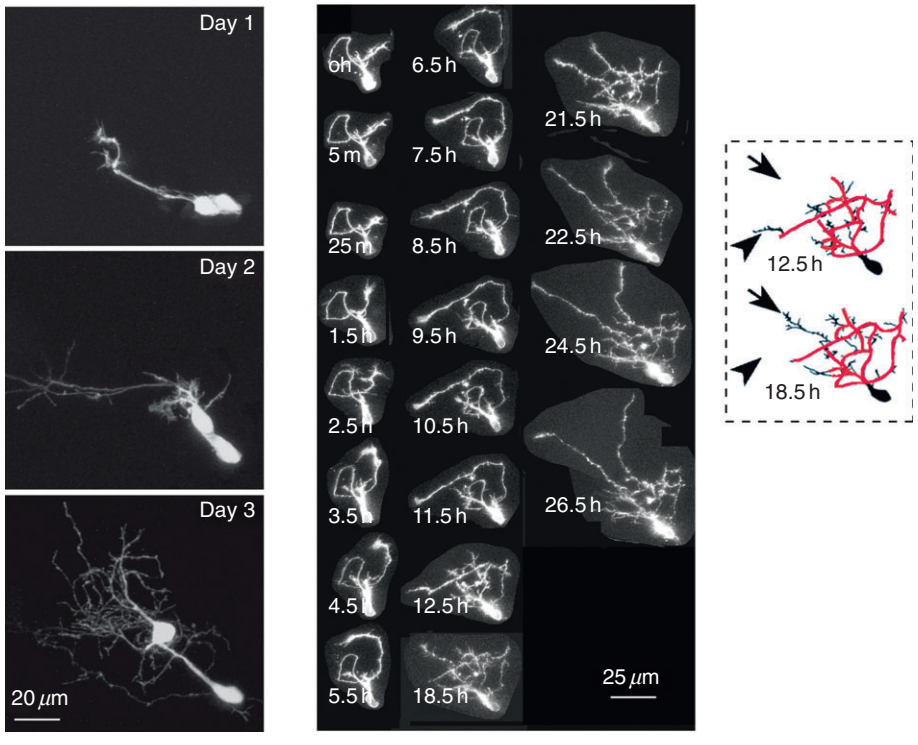


FIGURE 20.1 Time-lapse images of *Xenopus* optic tectal neurons collected *in vivo*. (A) This example shows two neighboring, newly differentiated neurons, close to the proliferative zone, imaged over a period of 3 days. These cells initially present glial-like morphologies (day 1). By the next day, these cells have elaborated complex dendritic arbors (day 2), which continue to grow to the end of imaging period (day 3). (B) Example showing an optic tectal interneuron imaged at short intervals over about 1 day. Dendritic trees develop as neurons differentiate within the tectum. Although portions of the dendritic arbor are stable over time (red outlined skeleton at 12.5 h and 18.5 h), other areas are very dynamic as dendritic branches are both added (arrows) and retracted (arrowheads) over time (in this case between 12.5 h and 18.5 h). *Source: From Bestman, Santos da Silva, and Cline (2008).*

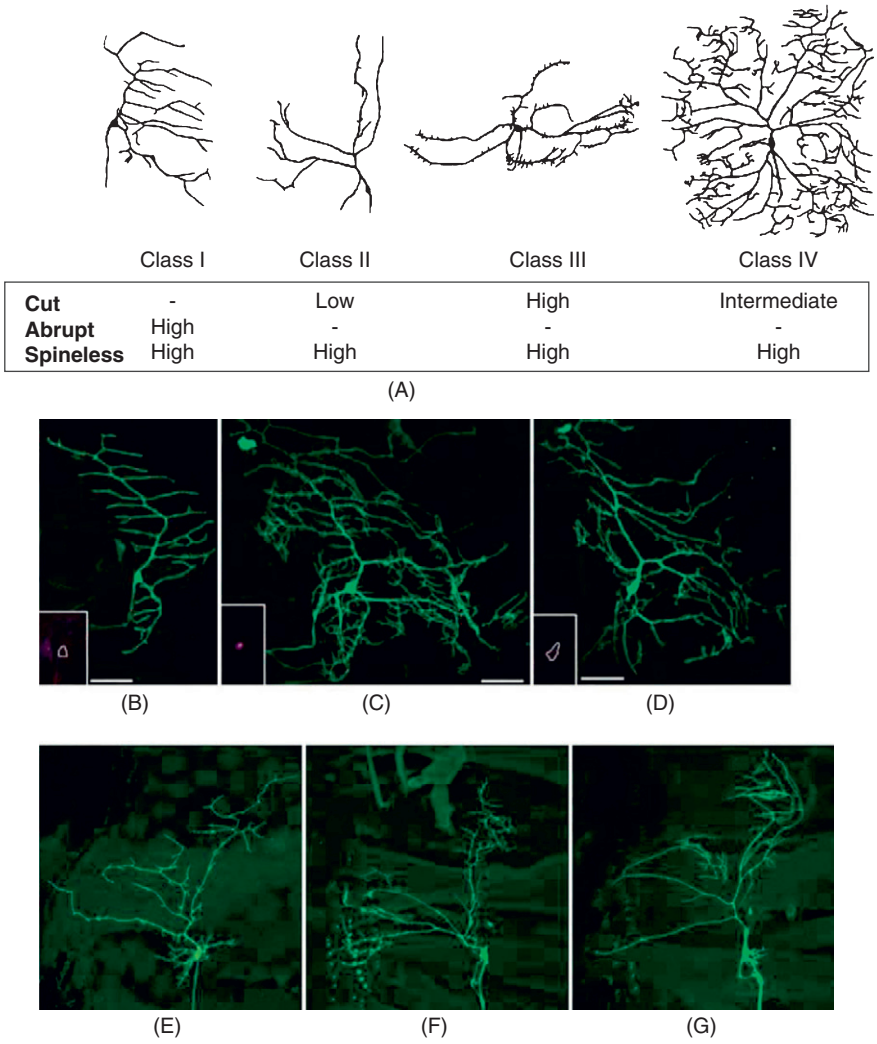


FIGURE 20.2 Transcription factors regulate the diversity and complexity of dendrites. (A) Dendrite morphologies of representative class I, II, III, and IV dendritic arborization (da) sensory neurons in the *Drosophila* PNS and a summary of the relative levels of expression of the transcription factors Cut, Abrupt, and Spineless in these neurons. (B–D). Ectopic expression of *cut* increases the dendritic complexity of class I da neurons. (B) Wild-type dendritic morphology of the ventral class I neuron vpda. Cut is normally not expressed in vpda (*inset*). (C) Ectopic expression of Cut in vpda leads to extensive dendritic outgrowth and branching. (D) Ectopic expression of CCAAT-displacement protein (CDP), a human homolog of *Drosophila cut*, also induces overbranching. (E–G) Loss of *spineless* function leads to a dramatic reduction in the dendritic diversity of different classes of da neurons. In loss-of-function *spineless* mutants, class I (E), class II (F), and class III (G) da neurons begin to resemble one another. *Source: From Parrish et al. (2007b).*

Transcription Factors Regulate Cell Type–Specific Dendritic Morphology

Expression of the gene *cut* in da neurons differs such that neurons with small and simple dendritic arbors either do not express Cut (class I neurons) or express low levels of Cut (class II), whereas neurons with more complex dendritic branching patterns and larger dendritic fields (class III and IV) express higher levels of Cut. Analysis of loss-of-function mutations and class-specific overexpression of Cut demonstrated that the level of Cut expression controls the distinct, class-specific patterns of dendritic branching (Grueber, Jan, & Jan, 2003a). Loss of Cut reduced dendrite growth and class-specific terminal branching and converted class III and IV neurons to class I and II morphologies with simpler dendritic branching patterns and smaller dendritic fields. Conversely, overexpression of Cut in neurons that express lower levels of endogenous Cut transformed branch morphology toward that seen in high-Cut expressing neurons. Furthermore, a human Cut homologue, CDP, can substitute for *Drosophila* Cut in promoting the dendritic morphology of high-Cut neurons (Fig. 20.2). Thus, Cut may function as an evolutionarily conserved regulator of neuronal-type specific dendrite morphologies.

Class III and class IV da neuron dendrites are made distinct by different combinations of Cut and another transcriptions factor Knot/Collier, which is expressed in Class IV but not in other classes of da neurons. In Class IV da neurons, Knot/Collier suppresses the tendency of Cut to induce actin-rich dendrite protrusions known as "dendritic spikes," structures unique to Class III da neurons. In contrast, Class III da neurons express high levels of Cut but not Knot/Collier—a combination that favors the formation of dendritic spikes. Thus, Knot/Collier and Cut provide an example of a combinatorial mechanism for specifying neuronal type–specific dendrite morphology (Jinushi-Nakao et al., 2007).

In contrast to Cut, Spineless (ss), the *Drosophila* homologue of the mammalian dioxin receptor, is expressed at

similar levels in all da neurons. In *ss* mutants, different classes of da neurons elaborate dendrites with similar branch numbers and complexities (Fig. 20.2), suggesting that da neurons might reside in a common "ground state" in the absence of *ss* function. Studies of the epistatic relationship between Cut and Spineless indicate that these transcription factors likely are acting in independent pathways to regulate morphogenesis of da neuron dendrites (Kim, Jan, & Jan, 2006). A comprehensive analysis of transcription factors with RNAi screens has revealed more than 70 transcription factors regulate dendritic arbor development of class I neurons in *Drosophila*. These findings suggest that complicated networks of transcriptional regulators likely regulate neuron-specific dendritic arborization patterns (Parrish, Emoto, Kim, & Jan, 2007b).

The Maintenance of Dendritic Fields

Dendrite development is a dynamic process involving both growth and retraction. Thus, selective stabilization or destabilization of branches might be one important mechanism to shape dendritic arbors. Studies of *Drosophila* class IV da neurons revealed that dendritic fields are actively maintained and there is a genetic program to maintain dendritic fields. The tumor suppressor Warts (Wts), as well as the Polycomb group of genes are required for the maintenance of the class IV da dendrites.

Drosophila has two NDR (nuclear Dbf2-related) families of kinase: Trc and Wts. Wts and its positive regulator Salvador originally were identified as tumor suppressor genes that function to coordinate cell proliferation and cell death. Loss-of-function mutants of either gene cause a progressive defect in the maintenance of the dendritic arbors, resulting in large gaps in the receptive fields (Fig. 20.3). Time-lapse studies suggest that the primary defect is in the maintenance of terminal dendrites, so Wts may normally function to stabilize these dendrites.

How are the establishment and maintenance of dendritic fields coordinated? In *Drosophila* class IV neurons, the Ste-20-related tumor suppressor kinase Hippo (Hpo) can directly phosphorylate and regulate both Trc, which functions in the establishment of dendritic tiling, and Wts, which functions in the maintenance of dendritic tiling (Emoto, Parrish, Jan, & Jan, 2006). Furthermore, *hpo* mutants have defects in both establishment and maintenance of dendritic fields. How Hpo regulates the transition from establishment to maintenance of dendritic fields remains to be determined.

What might be the downstream genes regulated by Wts? In the *Drosophila* retina, Wts regulates cell proliferation and apoptosis by phosphorylating the transcriptional coactivator Yorkie (Huang et al., 2005), but Yorkie does not appear to function in dendrite maintenance. Instead, the Polycomb genes are good candidates as targets for Wts/Sav for dendritic maintenance. The Polycomb genes regulate gene expression by repressing developmentally regulated genes. Polycomb genes can be separated into two multiprotein complexes: Polycomb repressor complex 1 (PRC 1) and PRC2. PRC2 is thought to mark the genes to be silenced by methylating histone H3, and PRC1 then comes in and blocks transcription. Mutants of several members of PRC1 and PRC2 have dendrite maintenance phenotypes very similar to that of Wts. Further, genetic and biochemical experiments suggest a functional link between Hpo/Wts signaling and the Polycomb genes, and that Polycomb genes regulate the dendritic field in part through Ultrabithorax (Ubx), one of the Hox genes in *Drosophila* (Parrish, Emoto, Jan, & Jan, 2007a).

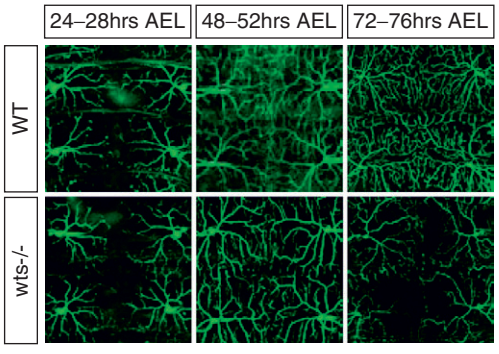


FIGURE 20.3 Dendritic fields are largely unchanged once established during development. Late-onset dendritic loss in *Drosophila warts* mutants (*wts*−/−) in late larval stages. Live images of wild-type (WT) and *wts* mutant (*wts*) dendrites of class IV da neurons at different times after egg laying (AEL). In *wts* mutants, dendrites initially tile the body wall normally but progressively lose branches at later larval stages. *Source: Adapted from Emoto et al. (2006).*

EXTRACELLULAR REGULATION OF DENDRITIC DEVELOPMENT

Regulation of Dendritic Differentiation and Dendrite Orientation

The developing cerebral cortex in mammals has been a valuable system to identify molecular mechanisms of dendritic growth control. Most excitatory cortical neurons are generated from precursors in the germinal zones lining the ventricle (Fig. 20.4). Once the cells become postmitotic, they migrate from the ventricular zone to the cortical plate. Following migration, pyramidal neurons extend an axon toward the ventricle and an apical dendrite toward the pial surface. To test the role of the local cortical environment in regulating the differentiation and guidance of nascent axons and dendrites, Polleux and Ghosh developed an *in vitro* assay in which dissociated neurons from a donor cortex were plated onto cortical slices in organotypic cultures. Neurons plated on cortical slices behave like the endogenous pyramidal neurons and

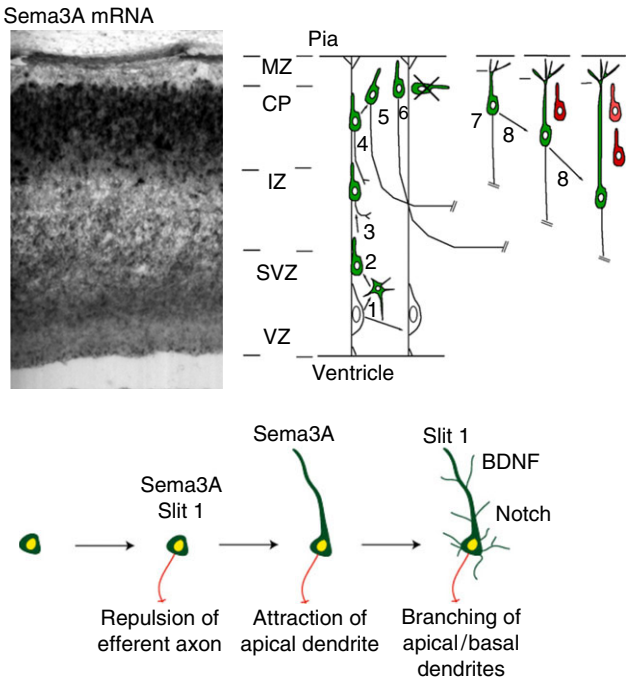


FIGURE 20.4 Upper Panel: Development of the dendritic morphology of cortical pyramidal neurons. Pyramidal neurons are generated from radial glial precursors in the dorsal telencephalon during embryonic development. Upon cell cycle exit from the ventricular zone (VZ), young post-mitotic neurons migrate along the radial glial scaffold and display a polarized morphology with a leading process directed toward the pial surface and sometimes a trailing process directed toward the ventricle. The leading process later becomes the apical dendrite. The trailing process of some neurons (but not all) develops into an axon that grows toward the intermediate zone (IZ; the future white matter) once cells reach the cortical plate (CP). Upon reaching the top of the cortical plate, postmitotic neurons detach from the radial glial processes and have to maintain their apical dendrite orientation toward the pial surface and axon outgrowth orientation toward the ventricle, which appears to be regulated by Sema3A, which acts as a chemoattractant for the apical dendrite and a chemorepellant for the axon. Adapted from Polleux and Ghosh (2008). Lower Panel: A model of how sequential action of extracellular factors might specify cortical neuron morphology. A newly postmitotic neuron arrives at the cortical plate, where it encounters a gradient of Sema3A (Polleux et al., 1998), which directs the growth of the axon towards the white matter. The same gradient of Sema3A attracts the apical dendrite of the neuron toward the pial surface (Polleux et al., 2000). Other factors, such as BDNF and Notch, control the subsequent growth and branching of dendrites. *Source: Adapted from (Polleux & Ghosh, 2008).*

extend an axon toward the ventricular zone and an apical dendrite toward the pial surface. Both the oriented growth of the axon and the apical dendrite are regulated by the chemotropic signal Sema 3A, which is present at high levels near the pial surface, and acts as a chemorepellant for axons and a chemoattractant for dendrites (Polleux, Giger, Ginty, Kolodkin, & Ghosh, 1998; Polleux, Morrow, & Ghosh, 2000).

The basis of the differential response of axons and dendrites to Sema3A appears to be asymmetric targeting of sGC to the emerging dendrite, which regulates cGMP levels (Polleux et al., 2000). Inhibition of cGMP-dependent kinase by pharmacological or genetic perturbation leads to ectopic dendritic orientation (Demyanenko et al., 2005; Polleux et al., 2000). Experiments carried out in *Xenopus* spinal neurons indicate that modulation of cGMP signaling by semaphorins can convert axons into dendrites, suggesting that local control of cGMP may be important in regulation of dendritic differentiation (Nishiyama et al., 2011). In addition to cGMP signaling, the nonreceptor tyrosine kinases Fyn and Cdk5 play important roles in mediating the effects of Sema3A on cortical dendrite orientation (Morita et al., 2006; Sasaki et al., 2002).

The concept of dendritic guidance turns out to be a general one: studies in *Drosophila* show that diffusible proteins (Netrin A and B) secreted by midline cells activate a cell-surface receptor (Frazzled) to pattern dendrites of motor neurons (Furrer, Kim, Wolf, & Chiba, 2003; Huber, Kolodkin, Ginty, & Cloutier, 2003; Yu & Bargmann, 2001).

Regulation of Dendritic Growth and Branching

The growth and branching of dendrites can be influenced by a large number of extracellular signals (see Fig. 20.9). In this section we discuss how specific extracellular factors regulate the development of the dendritic tree.

Neurotrophins

Neurotrophins (NGF, BDNF, NT-3, and NT-4) exert their effects on the growth and branching of dendritic arbors through the Trk family of tyrosine kinase receptors. Experiments in which the effects of neurotrophins on dendritic growth control have been examined in slice cultures from mammalian hippocampus or cortex indicate that, in general, neurotrophins increase dendritic complexity of pyramidal neurons by increasing total dendritic length, the number of branchpoints, and/or the number of primary dendrites (Baker, Dijkhuizen, Van Pelt, & Verhaagen, 1998; McAllister, Lo, & Katz, 1995; Niblock, Brunso-Bechtold, & Riddle, 2000; Yacoubian & Lo, 2000). By contrast, BDNF injection into *Xenopus* optic tectum increases synapse density on tectal neurons without affecting dendritic arbor structure (Sanchez et al., 2006).

The complexity of neuronal dendrites is likely to be influenced by the action of multiple neurotrophic factors. For example, Osteogenic Protein-1 (OP-1), a member of the transforming growth-factor-beta (TGFβ) superfamily, increases dendritic growth and branching in dissociated embryonic cortical neurons (Le Roux, Behar, Higgins, & Charette, 1999). Insulin-like growth factor-1 (IGF-1) affects dendrite growth and branching of postnatal layer 2 cortical neurons (Niblock et al., 2000). Furthermore, insulin receptor signaling is required for visual experience–dependent enhanced dendritic arbor growth in *Xenopus*

tectal neurons (Chiu, Chen, & Cline, 2008), illustrating that neurotrophin signaling mechanisms operate broadly across evolution.

Recent work indicates that Trk receptors and most other tyrosine kinase receptors affect dendrite development through the MAP kinase and PI-3 kinase pathways, which in turn include PTEN (phosphatase and tensin homolog deleted on chromosome Ten), mTOR (mammalian Target of Rapamycin), and local protein synthesis (Chow et al., 2009; Jaworski, Spangler, Seeburg, Hoogenraad, & Sheng, 2005). In addition to control of local protein synthesis, these signaling pathways appear to influence neuronal morphology by regulating the activity of the Rho family GTPases, which mediate actin cytoskeleton dynamics and are known to induce rapid dendritic remodeling (Box 20.1; Fig. 20.5).

Slit/Robo Signaling

The Slits are a well-studied family of multifunctional guidance cues that have been shown to both repel axons and migrating cells, as well as promote elongation and branching of developing sensory axons (reviewed in Huber et al., 2003). Generally, Slits exert their effects through binding to specific members of the Roundabout, or Robo, family of receptors. Slit1, and two of the three Robo receptors, Robo1 and Robo2, are expressed in the developing cortex during the time of initial axon and dendrite differentiation. Slit proteins simultaneously repel pyramidal neuron axons and stimulate dendrite growth and branching and these effects of Slit1 are mediated by the Robo1 and 2 receptors (Whitford et al., 2002). Thus, in contrast to the guidance role that Sema3A plays in orienting apical dendrites, Slit1 acts as a more general dendrite growth and branching signal for cortical neurons.

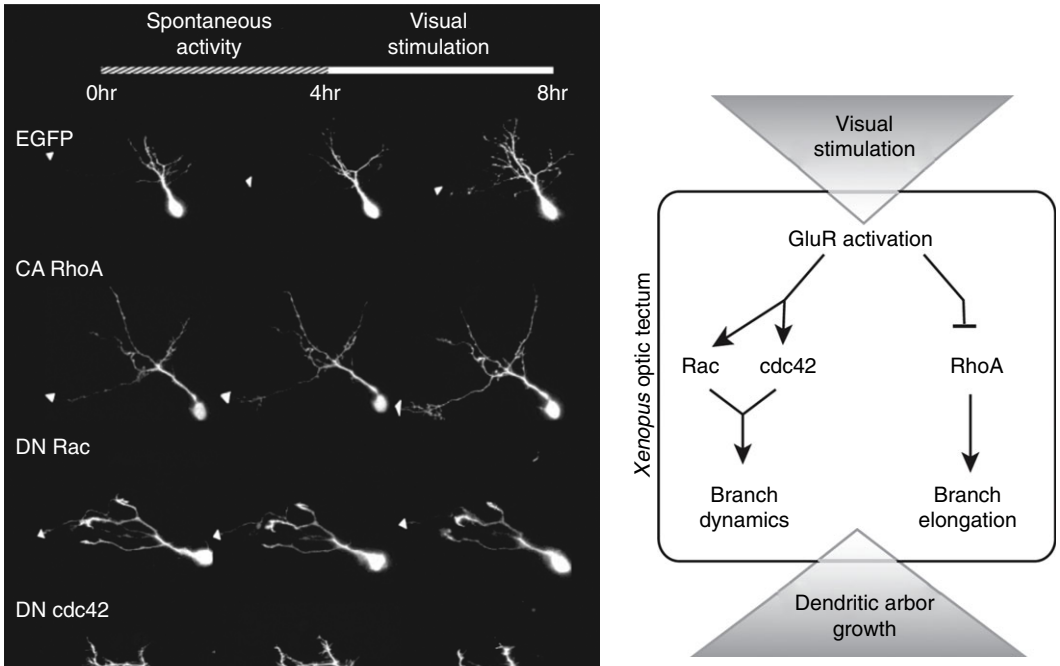
WNT Signaling

Another illustration of the ability of individual signals to control diverse biological responses including the control of axonal and dendritic development is provided by the WNT family of secreted proteins. The WNTs are a large family of proteins initially identified as potent morphogens involved in patterning organ development in both invertebrates and vertebrates. WNTs also have been shown to regulate cell proliferation, migration, and survival (Ciani, Krylova, Smalley, Dale, & Salinas, 2004; Ciani & Salinas, 2005). WNT proteins can function as axon-guidance molecules and as target-derived signals that regulate axonal remodeling and synapse formation (Budnik & Salinas, 2011; Hall, Lucas, & Salinas, 2000; Krylova et al., 2002). WNT proteins signal through at least three different pathways. The binding of WNT proteins to Frizzled receptors results in the activation of the scaffold protein Dishevelled (Dvl). In the so-called canonical pathway, WNT proteins signal through Dvl to inhibit GSK3-β a serine/threonine kinase. Inhibition of GSK3-β, in turn, activates β-catenin-mediated gene transcription. WNT proteins can also signal through Dvl to regulate Rho GTPases. Finally, WNT proteins can activate a Ca^2-dependent pathway, again through Dvl. WNT proteins induce axonal remodeling through the activation of Dvl and the subsequent inhibition of GSK3-β (Hall et al., 2000). Dvl has been shown to act locally to regulate microtubule stability by inhibiting a pool of GSK3-β through a β-catenin- and transcriptional-independent pathway (Ciani et al., 2004).

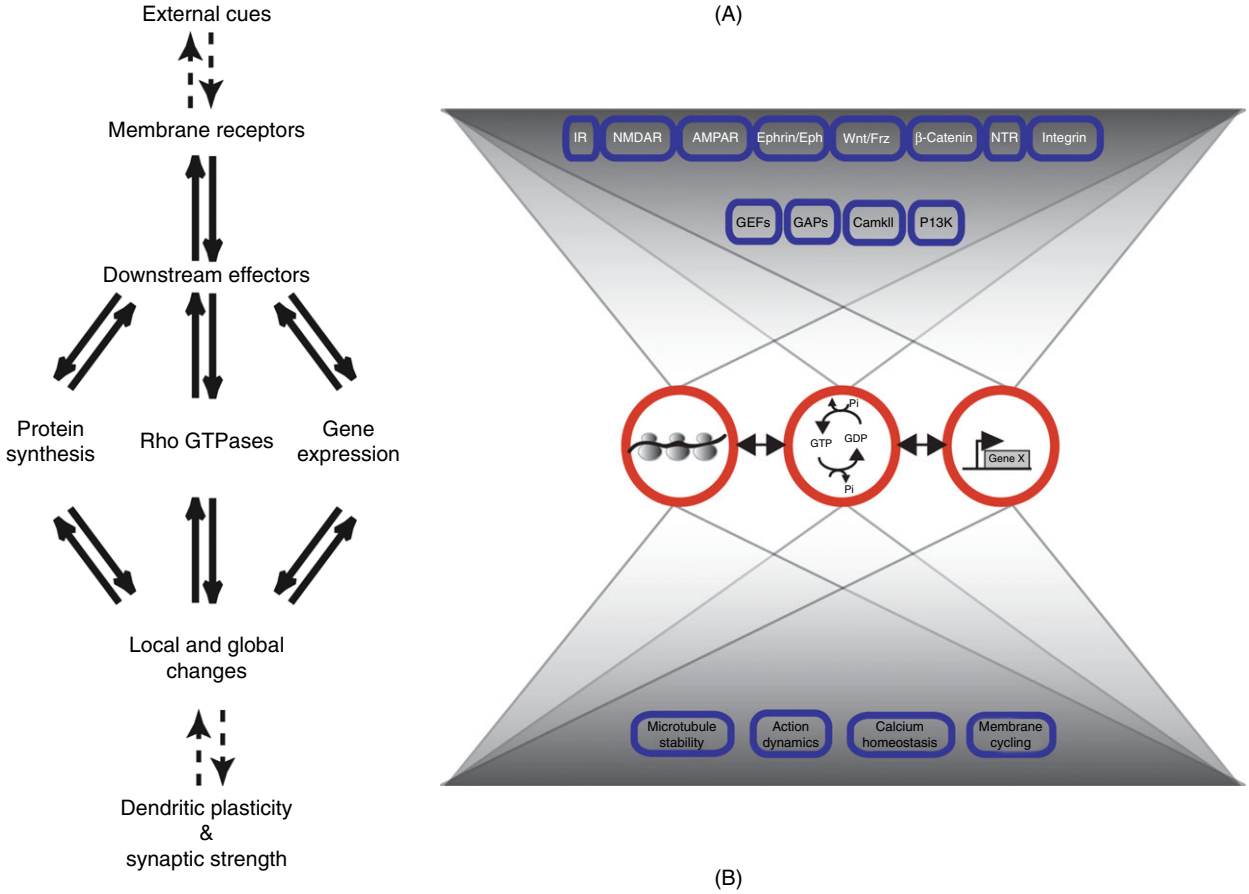
WNT proteins control dendritic arborization of hippocampal neurons (Rosso, Sussman, Wynshaw-Boris, & Salinas, 2005). Salinas and colleagues have found that *Wnt7b* is expressed in the mouse hippocampus and induces dendritic arborization of hippocampal neurons during development. This effect is mimicked by the expression of DVL. Importantly, analyses of the *Dvl1* mutant mouse revealed that DVL1 is crucial for dendrite development, as hippocampal neurons developed shorter and less complex dendrites in the mutant than in the wild-type. A dominant-negative β-catenin does not block DVL function in dendrites, suggesting that the WNT canonical pathway is not involved. In this case WNT7B and DVL signal through a noncanonical pathway in which the small GTPase Rac, but not Rho, is involved. First, endogenous DVL associates with Rac but not Rho in hippocampal neurons. Second, WNT7B or expression of DVL activates Rac in hippocampal neurons. Last, expression of a dominant- negative Rac blocks DVL function in dendrites.

Inhibition of the WNT pathway by SFRP1 (a secreted antagonist of WNT proteins) decreases Rac activation by WNT7B and blocks the effect of WNT7B in dendrite development (Rosso et al., 2005). Rosso et al. also report that WNT7B and DVL activate JNK, a downstream effector of Rac. Inhibition of JNK blocks DVL function in dendrites, whereas pharmacological activation of JNK enhances dendrite development. It remains to be determined whether JNK acts downstream of Rac, or whether the WNT–DVL pathway regulates Rac and JNK independently. Although Rho GTPases

FIGURE 20.5 Visual experience regulates dendritic growth through Rho, Rac, and Cdc42. (A) Visual stimulation over a 4-hour period increases the rate of dendrite growth in optic tectal neurons from *Xenopus*. Expression of constitutively active (CA) RhoA, dominant negative (DN) Rac, or DN Cdc42 affects specific aspects of dynamic dendritic arbor growth, as shown in the images of GFP-expressing tectal neurons collected *in vivo*. As summarized in the diagram on the right, visual stimulation, acting through glutamate receptors, triggers enhanced dendrite growth by regulating the Rho GTPases. (B) External cues acting through membrane receptors and downstream signaling pathways regulate functional and structural dendritic plasticity. Multiple mechanisms operate in parallel to accomplish and control neuronal plasticity. *Source: From Bestman et al. (2008).*


Spontaneous activity
Visual stimulation
0hr
4hr
8hr
EGFP
CA RhoA
DN Rac
DN cdc42
Visual stimulation
GluR activation
Xenopus optic tectum
Rac
cdc42
RhoA
Branch dynamics
Branch elongation
Dendritic arbor growth

(A)


External cues
Membrane receptors
Downstream effectors
Protein synthesis
Rho GTPases
Gene expression
Local and global changes
Dendritic plasticity & synaptic strength
IR
NMDAR
AMPAR
Ephrin/Eph
Wnt/Frz
β-Catenin
NTR
Integrin
GEFs
GAPs
CamkII
P13K
GTP
GDP
Pi
Gene X
Microtubule stability
Action dynamics
Calcium homeostasis
Membrane cycling

(B)

are well-known modulators of dendrite development and maintenance (see Box 20.1), the mechanisms by which extracellular factors modulate these molecules during dendrite morphogenesis remains poorly understood. The findings reported by Rosso et al. (2005) demonstrate that DVL functions as a link between WNT factors and Rho GTPases in dendrites, and they reveal a novel role for JNK in dendrite development.

Cadherins and Catenins

One of the central challenges in the study of dendritic development is to understand how extracellular cues that regulate dendritic branching are integrated with

BOX 20.1

RHO GTPASES CONTROL THE STRUCTURE OF THE DENDRITIC CYTOSKELETON

All the pathways that regulate dendritic development necessarily converge on the regulation of the cytoskeleton, which is controlled by the Rho GTPases. It is noteworthy that many, if not all, of these pathways also affect gene expression and different aspects of dendritic function, such as synaptic transmission, calcium signaling, and neuronal excitability. Such divergence of signaling from extracellular cues assures that the development of dendritic structure and function are tightly coregulated.

The dendritic cytoskeleton is composed of bundles of microtubules extending within the center of the dendritic shaft, a cortex of actin sandwiched between the microtubular bundles and the plasma membrane, and an actin matrix at the tip of dendritic processes. Fine terminal dendritic branches, or filopodia, have actin filaments as their sole cytoskeletal component (reviewed in Van Aelst and Cline (2004)). Considerable effort has been devoted to understanding the interaction between extracellular signaling events and the cytoskeleton, since these interactions are likely to be essential for the highly stereotyped and yet plastic elaboration of the dendritic arbor structure. A general scenario is emerging in which an extracellular signal interacts with a cell surface receptor that activates a cascade controlling the RhoA GTPases that, in turn, affect both the actin- and microtubule-based cytoskeleton in dendrites (Newey et al., 2005). The significance of these Rho GTPase regulatory molecules for dendritic morphogenesis is perhaps best illustrated by the fact that the abnormal development of dendritic trees, a hallmark of several different types of mental retardation, is at least in part caused by deficient signaling via Rho GTPases (Govek, Newey, & Van Aelst, 2005).

The Rho GTPases regulate the cytoskeleton in all cell types, but the elaborate and plastic structure of neurons poses particularly fascinating regulatory constraints on GTPase signaling (da Silva & Dotti, 2002; Luo, 2000; Van Aelst & Cline, 2004). The Rho GTPases function as bimodal switches, cycling between inactive, GDP-bound and active, GTP-bound conformations. RhoA, Rac1, and cdc42 are arguably the best studied of the small Rho GTPases. These molecules regulate both actin and microtubule dynamics (Gundersen, Gomes, & Wen, 2004; Zheng, 2004), and the manipulation of their individual activities has shown that each plays a particular role in dendritic structure development (reviewed in Newey et al. (2005)). The interplay of these effects on the cytoskeleton is key in shaping the intricacy of dendritic trees.

As more refined methods are used to probe the molecular and cellular basis of structural plasticity, our understanding of the intricate web of control becomes more complete. A striking example is that of the *in vivo* dendritic arbor development of optic tectal neurons in *Xenopus*. These neurons respond to stimulation of the tadpole visual system with an increased dendritic arbor growth rate, which requires glutamate receptor activity (Sin et al., 2002). Expression of dominant negative or constitutively active forms of RhoA, Rac, and Cdc42 demonstrated the participation of the Rho GTPases in the activity-dependent enhanced dendritic arbor growth rate. Expression of dominant negative forms of Rac or Cdc42, or expression of active RhoA blocked dendritic arbor elaboration in response to visual stimulation (Li, Aizenman, & Cline, 2002; Li, Van Aelst, & Cline, 2000; Sin et al., 2002) (Fig. 20.5A). These data suggest that glutamatergic synaptic input regulates the development of dendritic arbor structure by controlling cytoskeletal dynamics and that the Rho GTPases are an interface between glutamate receptor activity and the cytoskeleton. Furthermore, these data and reports from other systems support a model in which Rac and Cdc42 activity regulate rates of terminal branch dynamics, whereas RhoA regulates extension of branches in response to activity (Ahnert-Hilger et al., 2004; Hayashi, Ohshima, & Mikoshiba, 2002; Lee, Winter, Martícke, Lee, & Luo, 2000; Li et al., 2000, 2002; Nakayama, Harms, & Luo, 2000; Ng et al., 2002; Pilpel & Segal, 2004; Ruchhoeft, Ohnuma, McNeill, Holt, & Harris, 1999; Scott, Reuter, & Luo, 2003; Sin et al., 2002; Wong, Faulkner-Jones, Sanes, & Wong, 2000).

Branch formation requires the regulation of local cortical actin dynamics to create protrusive forces that allow filopodial sprouting (Luo, 2000). Time-lapse imaging indicates that filopodia are extremely dynamic, consistent with the

BOX 20.1 (*cont'd*)

rapid assembly and disassembly of actin filaments (Hossain, Hewapathirane, & Haas, 2011). The stabilization of filopodia and their extension as branches likely depend on their invasion by microtubules. Although the invasion of filopodia by microtubules is a key regulatory event in dendritic arbor development, the mechanisms regulating this process are unknown. Microtubules generate the mechanical forces necessary for branch elongation and can serve as tracks for the delivery of new membrane as branches extend (Horton & Ehlers, 2003, 2004), as well as transport of ribonucleoprotein (RNP) granules that regulate local protein synthesis (Bagni & Greenough, 2005). A close interplay between actin and microtubules is fundamental for the cellular events leading to changes in dendritic architecture. Importantly, in nonneuronal cells, the activity of Rho GTPases not only induces changes in both actin and microtubules but is also itself modified by alterations in the dynamics of the two cytoskeletons, thus serving as the regulator for the interplay between actin and microtubules (Etienne-Manneville, 2004; Fukata, Nakagawa, & Kaibuchi, 2003; Wittmann & Waterman-Storer, 2001; Zheng, 2004).

The role of RhoA in regulating the extension of branches in response to activity relies on its capacity to control microtubule stabilization (e.g., via mDia; Palazzo, Cook, Alberts, & Gundersen, 2001) and favor polymerization of cortical actin (Da Silva et al., 2003; Nobes & Hall, 1995), whereas Rac and cdc42 act on branch dynamics by regulating actin (e.g., supporting filopodial formation; Luo, 2002) and by favoring microtubular dynamics (e.g., by regulating catastrophe rates; Daub, Gevaert, Vandekerckhove, Sobel, & Hall, 2001; Kuntziger, Gavet, Manceau, Sobel, & Bornens, 2001). The description of the function of other Rho GTPases, such as that of Rnd2 in regulating branching via its effector Rapostlin (Negishi & Katoh, 2005), will help in the detailed understanding of how Rho GTPases regulate dendritic arbor development.

Rho GTPases are distributed ubiquitously throughout the neuronal cytoplasm (Govek et al., 2005), and, consequently, the activity of these proteins must be restrained in dendrites by resident upstream regulators. The link between incoming signals, for instance by activation of neurotransmitter receptors, and GTPase activity is mediated by GTPase regulatory proteins, which are particularly interesting because they are capable of integrating extracellular signaling with other signaling events relevant to neuronal structure. These include the guanine exchange factors (GEFs), which activate GTPases by favoring the substitution of GDP for GTP, and GTPase activating proteins (GAPs), which inactivate GTPases by inducing GTP hydrolysis (Bernards & Settleman, 2004; Schmidt & Hall, 2002). A rush of recent papers has examined the potential participation of several GEFs and GAPs in activity-dependent dendritic structural plasticity. GAPs can regulate dendritic development, as is exemplified by the observation that p190 RhoGAP is necessary for the dendritic remodeling that allows the shift from pyramidal to nonpyramidal morphologies in cortical cultures (Threadgill, Bobb, & Ghosh, 1997). Importantly, p190 RhoGAP is likely to exert its effect in an activity-dependent manner as indicated by its importance in fear memory formation in the lateral amygdala (Lamprecht, Farb, & LeDoux, 2002).

GEFs also play important roles in regulating dendrite arbor structure. For instance, Tiam1, a Rac-GEF, is located in dendrites and in particular in spines in cortical and hippocampal neurons. Tiam is noteworthy because it associates with the NMDA receptor, is phosphorylated in a calcium- and NMDA receptor–dependent manner, and is required for dendritic arbor development (Tolias et al., 2005). This is particularly interesting if one considers that the closely related member of the Dbl family of GEFs (Rossman & Sondek, 2005), Trio, regulates the development of axons in a potentially calcium-dependent manner (Debant et al., 1996), indicating the subcellular localization within different neuronal compartments is key to the specificity of GEF function. Kalirin, another example of a RhoGEF, in this case a dual RhoA- and Rac1-GEF, has been shown to regulate the development and maintenance of dendritic arbors by modulation of RhoA and Rac activities (Penzes et al., 2001). Recruitment of this Rho GTPase regulator in dendrites depends on the ephrin-EphB transynaptic signaling pathway, another cell surface signaling system linked to the actin cytoskeleton regulatory machinery. EphrinB-EphB receptor signaling may coordinate pre- and postsynaptic structural and functional development (Palmer & Klein, 2003). Its activation results in the translocation of Kalirin to synaptic sites and the activation of a signaling pathway involving Rac1 and the specific downstream effector PAK (Penzes et al., 2003). One intriguing possibility is that the Ephrin-EphR signaling could be coordinated with regulation of NMDA receptor distribution and calcium-permeability in postsynaptic sites (Dalva et al., 2000; Takasu, Dalva, Zigmond, & Greenberg, 2002). This kind of crosstalk between proteins involved in cell–cell contact and neurotransmitter receptors provides evidence for coregulation of development of dendritic arbor structure and synaptic communication.

Ca^2 activity-dependent signals. A potential clue comes from recent results exploring the role of β-catenin and cadherins in dendritic branching (Yu & Malenka, 2003). Overexpression of β-catenin (and other members of the cadherin/catenin complex) enhances dendritic arborization, whereas sequestering endogenous β-catenin causes a decrease in dendritic branching. Importantly, the authors show that blocking β-catenin prevents the enhancement of dendritic morphogenesis caused by neuronal depolarization. Furthermore, the release of secreted WNT, which occurs during normal neuronal development, is enhanced by manipulations that mimic increased activity, and WNTs contribute to the effects of neural activity on dendritic arborization. These results indicate that β-catenin is an important mediator of dendritic morphogenesis and that WNT β-catenin signaling is likely to be important during critical stages of dendritic development (Yu & Malenka, 2003).

These observations are reinforced by the recent demonstration that another class of cadherins (*Celsr1–3*) regulates dendritic branching of cortical pyramidal neurons (Shima, Kengaku, Hirano, Takeichi, & Uemura, 2004). This class of cadherins has been identified in vertebrates as orthologs of *flamingo*, a gene previously implicated in the control of dendritic development in *Drosophila* (Gao, Kohwi, Brenman, Jan, & Jan, 2000). Shima et al. (2004) combined loss-of-function techniques including RNAi-mediated gene silencing in single neurons using biolistic-delivery in P8 organotypic slice cultures to demonstrate that knocking-down *Celsr2* expression in both layer 5 pyramidal neurons and Purkinje cerebellar neurons significantly reduces dendritic branching (Shima et al., 2004). Furthermore, using the same technique, Shima et al. performed a structure-function analysis demonstrating that these effects of *Celsr2* on dendritic branching require the integrity of a site on the extracellular portion outside the cadherin-domain as well as a portion of the intracellular domain called the EGF-HRM region. These results indicate that cadherins play an important role in the control of dendritic complexity.

DENDRITIC DEVELOPMENT AND CIRCUIT FORMATION IN MAMMALIAN RETINA

Why are dendrites so elaborately patterned? A main reason is that their form underlies their function, and their function can be quite complex. The size, shape, branching pattern, and position of dendritic arbors are main determinants of what inputs they receive and how those inputs are integrated. These relationships have been demonstrated in many regions of the nervous system of both vertebrates and invertebrates. Here, to exemplify the relationship between dendritic development and circuit function, we focus on the vertebrate retina.

Retinal circuitry is described in Chapter 26, and summarized in Figure 26.2. Briefly, light excites photoreceptors, which form synapses on dendrites of bipolar cells; the axons of these cells form synapses on retinal ganglion cells (RGCs), whose axons send information to the brain. Horizontal cells and amacrine cells mediate lateral interactions. Complexity in this seemingly simple circuit arises from the existence of multiple subtypes of each main cell type—at least 12, 30, and 20 subtypes of bipolar, amacrine cells, and retinal ganglion cells, respectively, in mammals, along with several photoreceptor subtypes (red, blue, and green cones plus rods in humans) and 2–4 horizontal cell types in most species (but only one in mice). The synapses these cells make and receive are highly specific and the circuits they form are complex. As development proceeds, the dendrites of RGCs and amacrine cells and axons of bipolar cells extend to form the inner plexiform layer through a sequence of intercellular interactions (Godinho et al., 2005; Mumm et al., 2006). We will describe two main features of dendritic development, termed laminar and lateral specificity, that help explain how intrinsic and extrinsic factors regulate the development of retinal neuronal dendrites into circuits. We also mention studies from *Drosophila* da neurons, introduced above, in which similar phenomena have been studied in greater molecular depth.

LAMINAR SPECIFICITY IN THE INNER PLEXIFORM LAYER

The inner plexiform layer (IPL) is divided into a series of at least 10 narrow, parallel sublaminae, with arbors of individual RGCs confined to just one (monostratified) or two (bistratified) of the sublaminae. Distinct RGC subtypes display distinct sublaminar patterns. Likewise, axons of bipolar cells and processes of amacrine cells (amacrine cell processes both make and receive synapses but are generally called dendrites) are also confined to few sublaminae in subtype-specific patterns. Thus, the laminar restriction of dendrites and axons in the IPL is a main determinant of synaptic connectivity, and therefore the visual features to which subsets of RGC respond. For example, dendrites of RGCs that respond to decreases and increases in light intensity (so-called OFF and ON cells) are restricted to the outer and inner halves of the IPL, respectively, where they receive synapses from the specific subsets of interneurons that endow them with these response properties (Famiglietti & Kolb, 1976; Roska & Werblin, 2001).

How does laminar specificity arise? Most IPL synapses form prior to visual experience, though they may be modified by electrical activity (see below). An appealingly

simple idea is that expression of a homophilic adhesion molecule by pre- and postsynaptic partners, perhaps as a consequence of their cell autonomous differentiation programs, could match them to each other. In chick embryos, four closely related transmembrance recognition molecules of the immununoglobulin superfamily—Dscam, DscamL, Sidekick-1, and Sidekick-2—appear to play such roles. They are expressed by nonoverlapping subsets of bipolar, amacrine, and retinal ganglion cells that form synapses in distinct IPL sublaminae (Figure 20.6). Each of the four proteins is concentrated within the appropriate sublaminae and each mediates homophilic adhesion. Ectopic expression of any one of them in small groups of cells redirects their dendritic processes to the sublamina in which that immununoglobulin superfamily is most prominent. Conversely, when expression of any of the four is decreased, with interfering RNAs, the processes wander beyond their appropriate sublaminae (Yamagata & Sanes, 2008; 2010; Yamagata, Weiner, & Sanes, 2002). Together, these results suggest the existence of an "Ig superfamily code" for laminar specificity in retina. Genetic studies indicate that Dscams and Sidekicks also mediate laminar specificity in mouse retina (Fuerst, Harris, Johnson, & Burgess, 2010; Yamagata & Sanes, 2011), although the situation is complicated by the fact that Dscams (but not Sidekicks) also act as repellents implicated in dendritic patterning in the orthogonal plane (Fuerst, Koizumi, Masland, & Burgess, 2008; Fuerst et al., 2009; see below).

Recent studies have implicated the semaphorins in determining how particular connections come to occupy specific positions. The semaphorins are a diverse set of transmembrane and secreted ligands that signal through receptors of the plexin and neuropilin families. As noted above, the semaphorins had already been implicated in sculpting dendritic arbors of cortical neurons. Retinas from mice lacking various combinations of receptors and/or ligands have mistargetted arbors (Matsuoka et al., 2011a, 2011b). Clearly the codes have yet to be fully broken, but Dscams, Sidekicks, and Semaphorins account for only a fraction of the laminar choices that retinal axons make. However, an attractive model is that cues such as Semaphorins and Plexins help place arbors in appropriate laminar regions, where recognition molecules such as Sidekicks and Dscams further refine sublaminar patterns and promote specific connectivity among arbors that coexist within sublaminae.

Lateral Specificity in the Inner Plexiform Layer: Dendro–Dendritic Interactions Regulate the Shape and Organization of Dendritic Fields

Dendrites are highly organized within the "x-y plane" of the retina. This type of organization is very important for the function of the retinal circuit. For example, the size and shape of an RGC's arbor are prime determinants of its receptive field; the even spacing of neuronal dendrites and the degree of dendritic arbor overlap ensure uniform coverage of each part of the visual field by a full set of processing elements. Dendro–dendritic interactions have a profound influence on the size and shape of the dendritic field, as well as the spatial relationship between dendritic fields. In many areas of the nervous system, dendrites of different types of neurons are intermingled and packed into a tight space. This arrangement is not random but well organized. At least four mechanisms contribute to the orderly organization of dendritic fields: tiling, mosaicism, self-fasciculation, and self-avoidance (Fig. 20.7).

Tiling

Tiling, first discovered in mammalian retina (Wassle, Peichl, & Boycott, 1981), refers to an arrangement in which arbors of individual neurons of a single subtype overlap minimally. A leading model is that dendrites of a given neuronal class grow until they are stopped by repulsive signals from the dendrites of their neighbors, leading to an arrangement in which arbors efficiently fill a field while avoiding overlap (Grueber & Sagasti, 2010). The degree of tiling is quantified by a "coverage factor," which corresponds to the number of arbors of a given subtype a pin

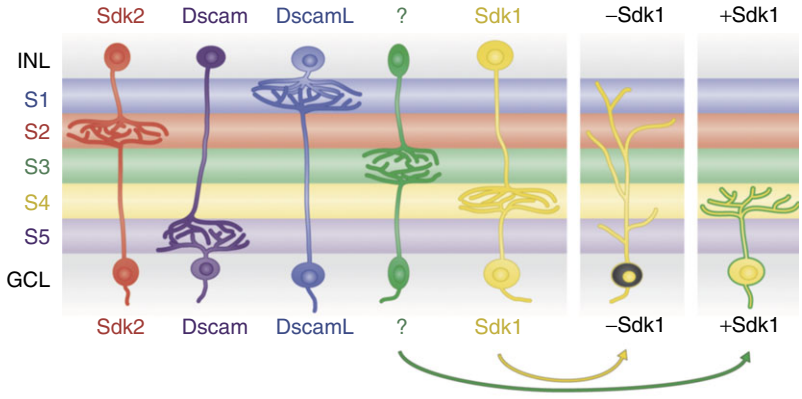


FIGURE 20.6 Expression of Sidekick 1, Sidekick 2, Dscam, and DscamL by nonoverlapping subsets of neurons in chick retina. Processes of most neurons that express each gene arborize in the same inner plexiform sublaminae, and altered expression of the gene (illustrated for Sidekick 1 in RGCs) leads to redistribution of their dendrites. *Source: From Sanes and Zipursky (2010).*

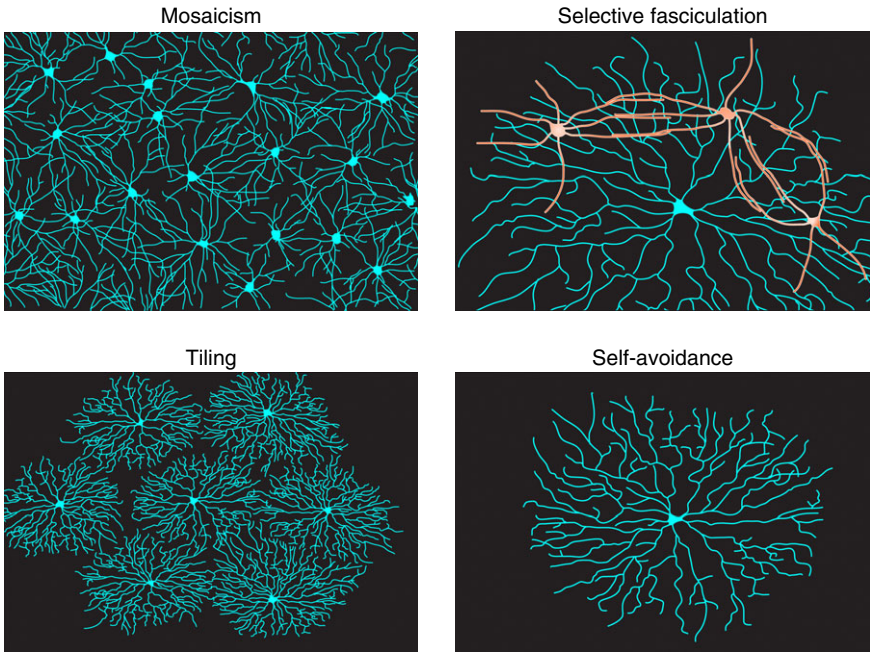


FIGURE 20.7 Four cellular processes that lead to regular arrangement of dendrites in the tangential plane of the retina's inner plexiform layer.

piercing the retina vertically would encounter. Thus, for perfect tiling, the coverage factor is 1.0. Dendrites of a few RGC subsets form tiled arrays in this sense (Dacey, 1993), as do the axonal arbors of many classes of bipolar cells (Wassle, Puller, Muller, & Haverkamp, 2009), but in fact dendrites of most RGC and amacrine subtypes do not tile: their coverage factors are significantly >1 and range up to >30 (Devries & Baylor, 1997; Peichl & Wassle, 1979). Thus, for these cells, regulation of dendritic size and control of dendritic overlap must arise from other mechanisms. One such mechanism is mosaicism (discussed below).

Tiling requires some cell surface recognition molecules to mediate the homotypic repulsion between neurons of the same type. Although the signal(s) that mediate vertebrate tiling behavior remain to be identified, studies in *Drosophila* and *C. elegans* have shown that the evolutionarily conserved protein kinase Tricornered (Trc) and the putative adaptor protein Furry (Fry), are important components of the intracellular signaling cascade involved in tiling (Emoto et al., 2004; Gallegos & Bargmann, 2004). In *trc* or *fry* mutants, dendrites no longer show their characteristic turning or retracting response when they encounter dendrites of the same type of neuron. In the mutants, unlike in the wild-type, there is extensive overlap of dendrites between adjacent neurons of the same kind. As a result, the mutant neurons have enlarged dendritic fields (Emoto et al., 2004).

Mosaicism

Neurons of most retinal subtypes are spaced more evenly than would be expected by chance alone. This regularity can be measured as the reduced probability of finding two cells of a single type near each other; the presence of an "exclusion zone" defines and characterizes the so-called "mosaic" arrangement (Eglen, 2006; Novelli, Resta, & Galli-Resta, 2005; Wassle & Riemann, 1978). Even in the absence of tiling, mosaics can help distribute each cell type evenly across the retina, thus optimizing dendritic coverage. Remarkably, mosaics are independent of each other; while a neuron of one subtype is unlikely to be adjacent to another of the same subtype, there is no restriction on its spatial relationship to neighboring neurons of other subtypes (Rockhill, Euler, & Masland, 2000), suggesting that molecular cues expressed by specific subtypes pattern mosaics by mediating homotypic (within-subtype) short-range repulsive interactions. Recently, a transmembrane protein called MEGF10 was found to mediate homotypic repulsive interactions critical for mosaic spacing of two interneuronal subtypes that occupy distinct cellular strata: horizontal cells and starburst amacrine cells (Kay & Sanes, 2011).

Selective Fasciculation

In some cases, dendrites of a particular neuronal subtype not only overlap, but they fasciculate. These associations presumably augment the specificity of the synapses that pass information laterally through the retina—for example, the electrical synapses that interconnect horizontal cells and the chemical synapses that interconnect starburst amacrine cells. Numerous cell adhesion molecules have been identified that promote fasciculation in other systems, and several are expressed by subsets of retinal neurons, but none has so far been shown to function in the IPL. On the other hand, the Dscams, mentioned above

as mediators of laminar specificity, appear to have a separate role as modulaters of fasciculation in mouse retina. Intriguingly, their role is a negative one; in their absence, neurites hyperfasciculate, leading to loss of selectivity and formation of large bundles that disrupt tiling and mosaicism (Fuerst et al., 2008, 2009). Thus, the selectivity of lateral interactions appears to require a proper balance between attractive and repellent interactions.

Self-Avoidance

Self-avoidance is a phenomenon in which neurites of a cell can recognize and avoid other neurites emanating from the same cell, while interacting freely with neurites emanating from other, similar neurons (Grueber & Sagasti, 2010). Self-avoidance thus leads to maximal and uniform coverage of receptive or projective fields by an individual neuron, while allowing multiple neurons to share the same field. The remarkable feature of self-avoidance is that it requires endowing neurons with individual molecular identities so they can discriminate between self and nonself neurites within a subtype (Zipursky & Sanes, 2010). For example, strict self-avoidance among retinal starburst amacrine cell dendrites is what endows them with the radial shape for which they are named, yet the dendrites bundle with dendrites of other starbursts and even form numerous synapses with them (Lefebvre & Sanes, 2011). Some RGC subtypes also show dendritic self-avoidance (Montague & Friedlander, 1991).

In *Drosophila*, da neuronal dendrites show self-avoidance that is attributable to the expression of Dscam1, a member of the immunoglobulin superfamily, originally identified as an axon guidance receptor. Alternative splicing of DSCAM1 in flies can potentially generate over 38,000 isoforms (Schmucker et al., 2000). A neuron typically expresses only a couple dozen of those isoforms (Neves, Zucker, Daly, & Chess, 2004). Dscam appears to be involved in self-avoidance through isoform-specific homophilic binding (Hummel et al., 2003; Wang, Zugates, Liang, Lee, & Lee, 2002; Zhu et al., 2006) in which neurites expressing the same set of Dscam isoforms repel one another (Wojtowicz et al., 2004). Mutant da neurons devoid of Dscam exhibit dendrite bundling and a crossing-over phenotype. This self-avoidance phenotype can be rescued by expressing a randomly selected single DSCAM isoform in the neuron, suggesting that Dscam is necessary in the da neuron for their dendritic self-avoidance, but the particular isoform is not important (Hughes et al., 2007; Matthews et al., 2007; Soba et al., 2007). In mice, Dscams may also play a role in self-avoidance, but they do not have numerous isoforms, so their roles are generic; they limit interactions of neurites both within an arbor and between arbors (isoneuronally and heteroneuronally). On the other hand, a gene cluster encoding a set of >50 protocadherins does exhibit isoform diversity, homophilic specificity, and combinatorial expression, suggesting that they could act in vertebrates as Dscam1 does in flies (Zipursky & Sanes, 2010). Indeed, deletion of the 22 genes of the protocadherin gamma subcluster markedly degrades self-avoidance among dendrites of mouse starburst amacrines (Lefebvre & Sanes, 2011).

ACTIVITY-DEPENDENT DENDRITIC DEVELOPMENT

Synaptic inputs play a significant role in controlling dendritic arbor development. Synaptic communication allows pre- and postsynaptic neurons to assess potential partners within a developing circuit, to establish and strengthen optimal connections, and to prune back or eliminate suboptimal contacts. Experience-dependent effects on brain development are readily demonstrated in sensory projections where afferent input can be manipulated from the periphery. *In vivo* imaging experiments in the visual system of *Xenopus* tadpoles showed that a relatively brief 4 hour period of simulated motion stimulus results in significant increased dendritic arbor growth rate, that requires glutamatergic synaptic transmission and Rho GTPase activity (Sin et al., 2002) (Fig. 20.5). Electrophysiological recordings from tectal neurons in animals with short-term visual enhancement (STVE) indicate that the visual experience increases synaptogenesis and synapse maturation (Aizenman & Cline, 2007), promotes maturation of tectal circuitry (Pratt, Dong, & Aizenman, 2008), and increases signal to noise in visual stimulus detection (Aizenman, Akerman, Jensen, & Cline, 2003). Recently experience-dependent increases in gene transcription were shown to regulate dendritic development and visual system function (Schwartz, Schohl, & Ruthazer, 2009, 2011). Together these studies indicate that sensory inputs activate a divergent range of cellular events that cooperate to orchestrate the development of functional circuits.

THE SYNAPTOTROPHIC MODEL OF DENDRITE DEVELOPMENT

How does activity regulate dendrite development? We will address this complex question by starting at the proximal site of activity-dependent regulation: the synapse. Considerable evidence supports the Synaptotrophic model of arbor development, presented by Vaughn (Cline & Haas, 2008; Vaughn, 1989; Vaughn, Barber, & Sims, 1988), which postulates that the formation and stabilization of synaptic contacts control the elaboration of dendritic (and axonal) arbors by regulating the stabilization or retraction of dynamic branches in the arbor (Fig. 20.8).

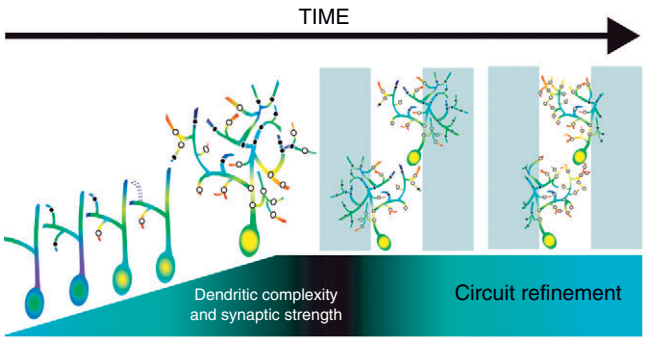


FIGURE 20.8 Schematic diagram of the effect of synaptic input on dendrite development. During dendritic development, newly generated glutamatergic synapses are mediated primarily by NMDA type glutamate receptors (black dots). With the addition of AMPA type glutamate receptors (white dots), synapses can be active at resting membrane conditions and overall synaptic strength is increased leading to dendritic branch stabilization. Dendrites with relatively weak synapses lacking AMPA receptors are retracted (dotted line). The resulting local increases in calcium influx through AMPA/NMDA containing synapses (indicated with "hot" colors) may enhance rates of branch addition and further stabilization. Newly stabilized branches become the substrate for further branch additions. It is the interplay between the dendrites and their synaptic partners that leads to selective stabilization and elaboration of branches toward appropriate target areas (gray and white bars) that is essential for circuit refinement. These processes are not just important during development, but underlie changes in circuit refinement in the mature nervous system. *Source: From Bestman et al. (2008).*

Although the greatest numerical increase in synapse density in mammalian cortex occurs after lamination has occurred and many pyramidal cells have extended dendrites (Rakic, Bourgeois, Eckenhoff, Zecevic, & Goldman-Rakic, 1986), a key observation from Vaughn's EM studies was that synapses form on arbors as the arbors elaborate (Vaughn, 1989; Vaughn et al., 1988). Indeed, whole cell recordings from newly differentiated optic tectal neurons in *Xenopus* or cortical neurons in turtle demonstrate that they have glutamatergic and GABA ergic synaptic inputs as soon as they extend their first dendrites (Blanton & Kriegstein, 1991; Wu, Malinow, & Cline, 1996).

Glutamatergic synapse formation is a multistep process in which an initial adhesion event is rapidly followed by synaptic transmission mediated by NMDA receptors. NMDA receptor activity promotes the trafficking of AMPA receptors into nascent synapses, which renders the synapses functional of resting potential (Cline, Wu, & Malinow, 1997). The Synaptotrophic hypothesis predicts that manipulations of synapse formation or synapse stabilization would affect dendritic arbor development. Adhesive events mediated by neurexin-neuroligin were recently shown to be required for tectal cell dendritic development (Chen, Tari, She, & Haas, 2010), consistent with Vaughn's idea. Time-lapse *in vivo* imaging experiments showed that blockade of NMDA or AMPA receptors or interference with their trafficking perturbs tectal cell dendritic arbor development by interfering with addition or stabilization of dendritic branches (Haas, Li, & Cline, 2006; Rajan & Cline, 1998). Genetically decreasing NMDA receptor subunit expression also disrupted dendritic arbor development in the rodent somatosensory system (Lee, Lo, & Erzurumlu, 2005). Time-lapse imaging demonstrated that blocking trafficking of AMPA receptors into developing synapses increased the rates of branch retractions in the developing dendritic arbor and increased the proportion of transient dendritic branches (Cline & Haas, 2008; Haas et al., 2006), exactly as predicted by the Synaptotrophic hypothesis, that strengthening synapses formed onto newly established dendritic branches would stabilize the branches. Similarly in the hippocampus and cortex, perturbing AMPA receptor activation or trafficking affects synapse maturation and dendrite arborization on pyramidal neurons (De Marco Garcia, Karayannis, & Fishell, 2011; Shi, Cheng, Jan, & Jan, 2004a; Shi, Cox, Wang, Jan, & Jan, 2004b), again supporting the general principle that afferent inputs control dendritic arbor development *in vivo*.

During early stages of neuronal development, GABAergic synaptic activity depolarizes neurons from their resting potential because the reversal potential for chloride is relatively positive due to the delayed expression of the KCC2 chloride transporter. In *Xenopus* neurons, premature expression of KCC2 blocked the maturation of glutamatergic synaptic inputs and indicated that depolarizing GABAergic inputs cooperate with NMDA receptors to promote AMPA receptor trafficking into developing synapses (Akerman & Cline, 2006). Premature expression of KCC2 in cortical neurons impaired dendritic arbor development (Cancedda, Fiumelli, Chen, & Poo, 2007), suggesting that depolarizing GABAergic input might affect dendritic arbor development indirectly through an effect on development of excitatory synaptic inputs. It is interesting to note that once dendritic arbors are elaborated, glutamatergic synaptic input activity and CaMKII stabilize dendritic arbor structure (Wu & Cline, 1998), suggesting that signaling cascades downstream of these events proximal to the synapse change as the neuron and circuit develop.

Recent studies provided the first evidence that inhibitory GABAergic inputs regulate dendritic arbor development. Shen, Da Silva, He, and Cline (2009) expressed a peptide in optic tectal neurons, called ICL, which interfered with the residence of $GABA_A$ receptors at synaptic sites. ICL expression selectively decreased inhibitory GABAergic, but not glutamatergic, synaptic inputs onto tectal neurons, shown by electrophysiological criteria. *In vivo* imaging experiments showed that ICL expression severely disrupted dendritic arbor development and experience-dependent structural plasticity, in striking similarity to the effect of interfering with glutamatergic

synaptic maturation (Haas et al., 2006). Further studies showed that ICL-expression in tectal neurons significantly disrupted visual information processing and visually guided behavior (Shen, McKeown, Demas, & Cline, 2011). These studies demonstrate that both excitatory and inhibitory synaptic inputs regulate dendritic arbor development, but whether similar or unique mechanisms operate under these circumstances is not yet known.

BRANCH RETRACTION AND SYNAPSE ELIMINATION

Although dendrite arbor elaboration clearly requires net branch addition and extension, both branch and synapse elimination are integral parts of dendritic arbor development. The final structure of cortical pyramidal dendrites is heavily influenced by pruning major portions of the arbor. For instance, a recent study in ferret visual cortex showed that layer 4 neurons initially form an apical dendrite, which is subsequently retracted to produce the classical stellate morphology. Furthermore, visual input is required for the pruning of the apical process (Callaway & Borrell, 2011). Another clear example of this phenomenon comes from study of the retina, where developing retinal ganglion cells transiently respond to both light on and light off events and their dendritic arbors terminate in both On and Off neuropil laminae. With normal visual experience, the bistratified dendritic arbors are pruned to On and Off laminae. Dark-rearing blocks the normal pruning of dendritic branches in retinal ganglion cells so that retinal ganglion cells remain responsive to On and Off visual stimuli (Tian & Copenhagen, 2003). In contrast to the relatively large-scale dendrite pruning mentioned above, a recent study, which combined in time-lapse imaging and subsequent serial section electron microscopic reconstruction of the imaged neuron, showed that synapse elimination is a prominent activity-dependent component of microcircuit assembly (Li, Erisir, & Cline, 2011). In this study, time-lapse *in vivo* imaging over periods of hours or days allowed identification of newly added or relatively stable dendritic branches, while the serial section EM allowed ultrastructural comparison of synaptic contacts on identified dynamic and stable branches. Although one might anticipate that new dendritic branches might have few synapses and that synapse density might increase with branch stabilization, this study showed otherwise. Surprising, newly emerged dendritic branches had a significantly higher synapse density compared to more stable branches. As dendritic branches stabilize, most synapses are eliminated and one input remains and becomes morphologically mature (Li & Cline, 2010). Furthermore, visual experience and NMDA receptor activity were required for both synapse elimination and synapse maturation. These data provide quantitative ultrastructural understanding of the activity-regulated cellular mechanisms contributing to dendritic arbor development.

CALCIUM-DEPENDENT MECHANISMS THAT MEDIATE DENDRITIC GROWTH

The effects of neuronal activity on dendritic arbor development are mediated by calcium signaling. Calcium levels in neurons are regulated by influx through voltage-sensitive calcium channels (VSCC), NMDA receptors, and calcium-permeable AMPA receptors, as well as by release of calcium from intracellular stores (Fig. 20.9). Release from internal stores principally involves calcium-induced calcium release (CICR) or activation by ligands that lead to the production of IP_3, which acts on internal stores. Two major signaling targets of elevated intracellular calcium are calcium/calmodulin dependent protein kinases (CaMKs) and mitogen-activated kinase (MAPK), which have multiple effects on dendrite development that reflect rapid and local consequences of kinase activity, as well as regulation of gene transcription through calcium-responsive transcription factors, such as CREB and CREST.

Of the CaMKs, CaMKII has been most extensively studied in relation to a role in dendritic development and function because both CaMKII isoforms, αCaMKII and βCaMKII, are present at high levels in dendrites. Most functional heteromers of CAMKII in cells include both α and β isoforms, and the α to β ratio in the heteromer regulates its association with the actin cytoskeleton and therefore its microdistribution in cells (Lisman, Schulman, & Cline, 2002). CaMKII strengthens synapses and stabilizes or restricts dendritic growth of frog tectal neurons *in vivo* and mammalian cortical neurons *in vitro* (Redmond, Oh, Hicks, Weinmaster, & Ghosh, 2000; Wu & Cline, 1998). CaMKIV is predominately localized in the nucleus, where it can respond to cytosolic calcium transients that enter the nucleus and regulate dendritic morphology through transcription events (Mauceri, Freitag, Oliveira, Bengtson, & Bading, 2011). The best-characterized target of CaMKIV is the transcription factor CREB. A noteworthy effector of CREB is CPG15, candidate plasticity gene 15 (Nedivi, 1999; Nedivi, Fieldust, Theill, & Hevron, 1996; Nedivi, Hevroni, Naot, Israeli, & Citri, 1993). CPG15 is a highly conserved neurotrophic molecule, which can promote neural survival and differentiation in mammalian systems (Putz, Harwell, & Nedivi, 2005). CPG15 enhances dendritic and axon arbor elaboration, and increases synaptogenesis and synaptic strength when expressed in *Xenopus* optic tectal cells *in vivo* (Cantallops, Haas, & Cline, 2000; Nedivi et al., 1998). Furthermore, the trafficking of GPI-linked CPG15 to the surface of axons is activity-regulated (Cantallops & Cline, 2008), suggesting that

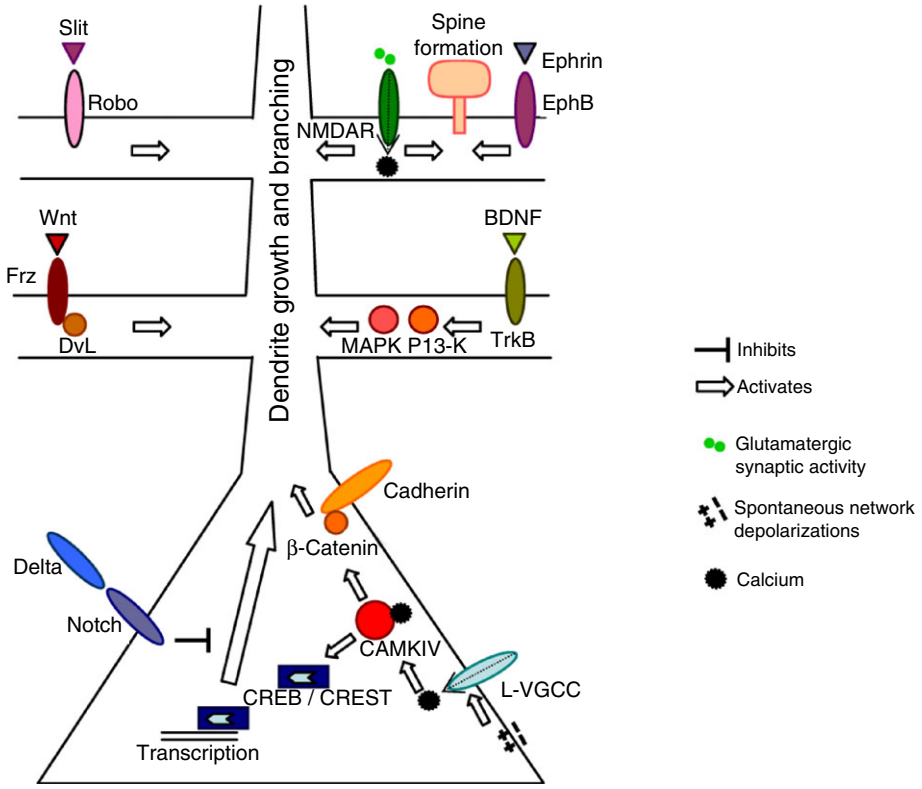


FIGURE 20.9 A summary of signaling pathways by which neuronal activity influences dendritic development. The effects of neuronal activity on dendritic development are mediated by calcium influx via voltage-gated calcium channels (VGCCs) and NMDA receptors, as well as release from internal stores. Local calcium signals act via Rho family proteins to regulate dendritic branch dynamics and stability, whereas global calcium signals recruit transcriptional mechanisms to regulate dendritic growth.

sensory activity regulates both the transcription and cellular localization of CPG15, which in turn promotes and coordinates multiple aspects of circuit development.

PROTEIN SYNTHESIS DEPENDENT REGULATION OF DENDRITE DEVELOPMENT

Protein synthesis can be regulated locally within dendrites by mRNA binding proteins, such as the fragile X mental retardation protein, FMRP, cytoplasmic polyadenylation element binding protein (CPEB), staufen, and pumilio. In general, these proteins bind specific mRNA cargo and hold them translationally silent until they are signaled to release the cargo for translational processing. The possibility that local protein synthesis in dendrites might affect neuronal morphology was first suggested from studies on material from patients with Fragile X Syndrome (Purpura, 1974), in which the spines on pyramidal cell dendrites were unusually long and thin. Work in a variety of systems has begun to elucidate the role of mRNA binding proteins in dendrite development. For instance, in *Xenopus*, *in vivo* imaging of neurons expressing dominant negative forms of CPEB showed that CPEB RNA binding and microtubule-based transport into dendrites was required for normal dendritic arbor development and visual experience–dependent structural plasticity. Furthermore, electrophysiological recordings show that cells that are deficient in CPEB function fail to integrate into the retinotectal circuit and do not respond to visual inputs (Bestman & Cline, 2008). Work in rodent visual cortex indicated that CaMKII translation was regulated by CPEB1 downstream of visual experience and NMDA receptor activity (Wells et al., 2001). Many candidate mRNA cargoes of CPEB have been identified bioinformatically. Given that CaMKII, which is just one protein that is translationally regulated by CPEB, is itself pleiotropic with respect to control of dendrite arbor development, it is clear that we must embrace the complexity of activity-dependent regulatory mechanisms controlling dendritic arbor development and circuit connectivity.

CONVERGENCE AND DIVERGENCE

A wide variety of signaling pathways participate in dendritic arbor development by regulating different downstream events such as cytoskeletal rearrangements, Ca homeostasis, and membrane cycling (Fig. 20.10). In addition, many signaling cascades diverge, for example to regulate somatic gene expression, local protein synthesis as well as the Rho GTPases. These diverging pathways ultimately result in modifications of dendritic plasticity. One example

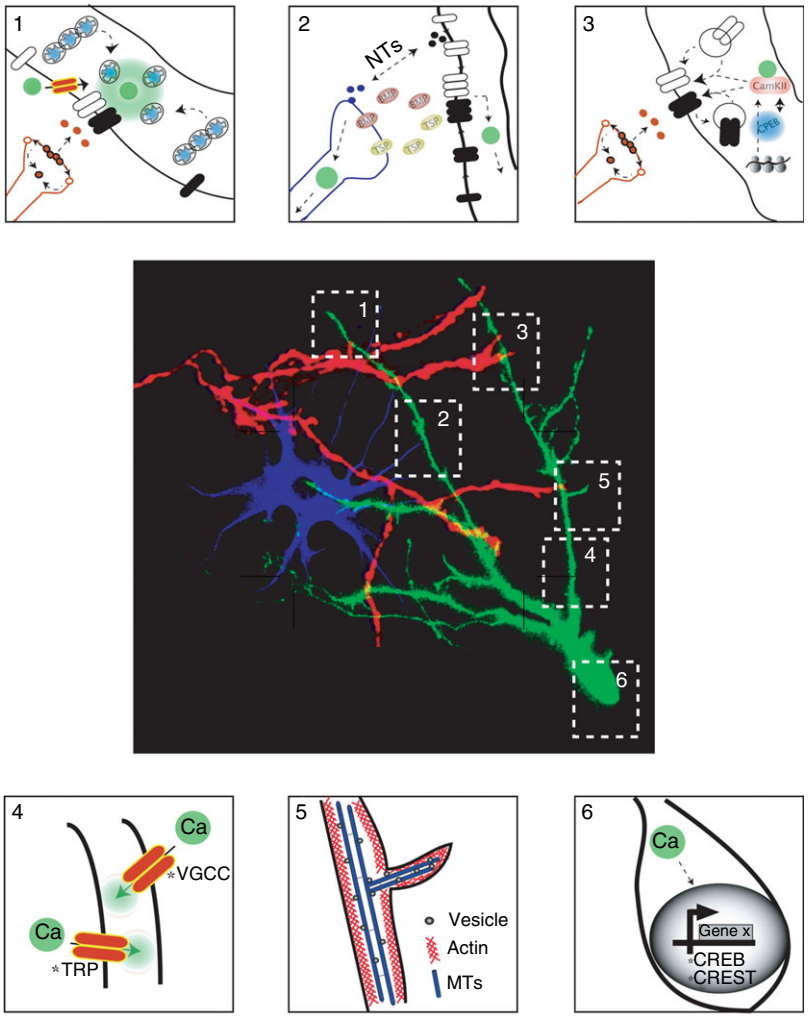


FIGURE 20.10 Summary of multiple mechanisms that influence dendritic growth and remodeling. Dendritic arbors develop within a complex environment in which they contact and receive signals from afferent axons, glial cells, and other dendrites. Several changes occur at sites of contact between axons and dendrites, marked by 1 and 3 in the image, including local changes in enzyme activity, such as CaM kinase and phosphatases, receptor trafficking, and local protein synthesis. Interactions between glia and neurons, marked by #2, include release of trophic substances that regulate synapse formation and maintenance. Process outgrowth, represented by #4 and #5, is mediated by cytoskeletal rearrangements. Finally activity-induced gene transcription, #6, can change the constellation of protein components in neurons in response to growth factors or synaptic inputs.

of divergence comes from the activation of the Ras superfamily GTPases that regulate membrane cycling downstream of neurotransmitter receptor activity. Synaptic NMDA receptors activate the Rab GTPase Rab5 (Pfeffer & Aivazian, 2004) that in turn regulates the internalization and dephosphorylation of AMPA receptors (Brown, Tran, Backos, & Esteban, 2005), an event associated with changes in dendritic structure (Ikegaya et al., 2001) and synaptic strength. Similarly, decreasing or increasing the activity of the Ras GTPase Rap1 decreases or increases, respectively, dendritic arbor elaboration *in vitro* (Chen, Wang, & Ghosh, 2005). One way Rap1 works is by regulating how calcium influx affects CREB-dependent transcription (Chen et al., 2005). Rap1 also mediates the internalization of AMPA type glutamate receptors from local synaptic sites in response to NMDA receptor signaling (Zhu, Qin, Zhao, Van Aelst, & Malinow, 2002). Since blocking glutamate receptor activity also decreases dendritic arbor development (Rajan & Cline, 1998), it is possible that Rap1 may act principally to control cell surface levels of glutamate receptors and this, in turn, affects dendritic arbor elaboration, as shown by interfering with AMPA receptor trafficking (Haas et al., 2006; Shi et al., 2004b). As with the Rho GTPases, Rap1 and the other GTPases regulating membrane cycling are at nodal points in the cascades controlling cell surface distribution of membrane proteins. For instance, Rap1 is downstream of other receptors, such as the neurotrophin receptors, plus GTPase-dependent membrane cycling regulates the cell surface distribution of signaling molecules, other than glutamate receptors. As illustrated in Figure 20.10, the pathways regulating dendritic arbor development are nonlinear and neither their regulation nor their targets are independent. Consequently, it is essential to investigate the function of these complex molecules in their normal *in vivo* context in order to dissect their endogenous functions.

CONCLUSION

Studies such as those summarized here, aided by computational approaches to map biochemical events in space and time, will ultimately lead to a complete appreciation

of the signaling pathways that regulate dendritic structural and functional plasticity. Elucidation of these pathways will reveal mechanisms by which activity can rapidly affect dendritic structure, and over more-prolonged time frames, exert profound changes in the so-called intrinsic states of the dendritic tree by modifying gene expression (Goldberg, 2004). Knowledge of these events, together with an understanding of the integrative capacity of dendrites, is essential to grasp the function of neuronal circuits and the brain.

References

Ahnert-Hilger, G., Holtje, M., Grosse, G., Pickert, G., Mucke, C., Nixdorf-Bergweiler, B., & Just, I. (2004). Differential effects of Rho GTPases on axonal and dendritic development in hippocampal neurones. *Journal of Neurochemistry, 90,* 9–18.

Aizenman, C. D., Akerman, C. J., Jensen, K. R., & Cline, H. T. (2003). Visually driven regulation of intrinsic neuronal excitability improves stimulus detection *in vivo*. *Neuron, 39,* 831–842.

Aizenman, C. D., & Cline, H. T. (2007). Enhanced visual activity *in vivo* forms nascent synapses in the developing retinotectal projection. *Journal of Neurophysiology, 97,* 2949–2957.

Akerman, C. J., & Cline, H. T. (2006). Depolarizing GABAergic conductances regulate the balance of excitation to inhibition in the developing retinotectal circuit *in vivo*. *The Journal of Neuroscience, 26,* 5117–5130.

Bagni, C., & Greenough, W. T. (2005). From mRNP trafficking to spine dysmorphogenesis: The roots of fragile X syndrome. *Nature Reviews. Neuroscience, 6,* 376–387.

Baker, R. E., Dijkhuizen, P. A., Van Pelt, J., & Verhaagen, J. (1998). Growth of pyramidal, but not non-pyramidal, dendrites in long-term organotypic explants of neonatal rat neocortex chronically exposed to neurotrophin-3. *The European Journal of Neuroscience, 10,* 1037–1044.

Bernards, A., & Settleman, J. (2004). GAP control: Regulating the regulators of small GTPases. *Trends in Cell Biology, 14,* 377–385.

Bestman, J. E., & Cline, H. T. (2008). The RNA binding protein CPEB regulates dendrite morphogenesis and neuronal circuit assembly *in vivo*. *Proceedings of the National Academy of Sciences of the United States of America, 105,* 20494–20499.

Bestman, J., Santos Da Silva, J., & Cline, H. T. (2008). Dendrite development. In *Dendrites*, G. Stuart, N. Spruston, & M. Hausser (Eds.). Oxford, New York: Oxford University Press.

Blanton, M. G., & Kriegstein, A. R. (1991). Spontaneous action potential activity and synaptic currents in the embryonic turtle cerebral cortex. *The Journal of Neuroscience*, 3907–3923.

Brown, T. C., Tran, I. C., Backos, D. S., & Esteban, J. A. (2005). NMDA receptor-dependent activation of the small GTPase Rab5 drives the removal of synaptic AMPA receptors during hippocampal LTD. *Neuron, 45,* 81–94.

Budnik, V., & Salinas, P. C. (2011). Wnt signaling during synaptic development and plasticity. *Current Opinion in Neurobiology, 21,* 151–159.

Callaway, E. M., & Borrell, V. (2011). Developmental sculpting of dendritic morphology of layer 4 neurons in visual cortex: Influence of retinal input. *The Journal of Neuroscience, 31,* 7456–7470.

Cancedda, L., Fiumelli, H., Chen, K., & Poo, M. M. (2007). Excitatory GABA action is essential for morphological maturation of cortical neurons *in vivo*. *The Journal of Neuroscience, 27,* 5224–5235.

Cantallops, I., & Cline, H. T. (2008). Rapid activity-dependent delivery of the neurotrophic protein CPG15 to the axon surface of neurons in intact Xenopus tadpoles. *Developmental Neurobiology, 68,* 744–759.

Cantallops, I., Haas, K., & Cline, H. T. (2000). Postsynaptic CPG15 promotes synaptic maturation and presynaptic axon arbor elaboration *in vivo*. *Nature Neuroscience, 3,* 1004–1011.

Chen, S. X., Tari, P. K., She, K., & Haas, K. (2010). Neurexin-neuroligin cell adhesion complexes contribute to synaptotropic dendritogenesis via growth stabilization mechanisms *in vivo*. *Neuron, 67,* 967–983.

Chen, Y., Wang, P. Y., & Ghosh, A. (2005). Regulation of cortical dendrite development by Rap1 signaling. *Molecular and Cellular Neurosciences, 28,* 215–228.

Chiu, S. L., Chen, C. M., & Cline, H. T. (2008). Insulin receptor signaling regulates synapse number, dendritic plasticity, and circuit function *in vivo*. *Neuron, 58,* 708–719.

Chow, D. K., Groszer, M., Pribadi, M., Machniki, M., Carmichael, S. T., Liu, X., & Trachtenberg, J. T. (2009). Laminar and compartmental regulation of dendritic growth in mature cortex. *Nature Neuroscience, 12,* 116–118.

Ciani, L., Krylova, O., Smalley, M. J., Dale, T. C., & Salinas, P. C. (2004). A divergent canonical WNT-signaling pathway regulates microtubule dynamics: Dishevelled signals locally to stabilize microtubules. *The Journal of Cell Biology, 164,* 243–253.

Ciani, L., & Salinas, P. C. (2005). WNTs in the vertebrate nervous system: From patterning to neuronal connectivity. *Nature Reviews. Neuroscience, 6,* 351–362.

Cline, H. T. (2001). Dendritic arbor development and synaptogenesis. *Current Opinion in Neurobiology, 11,* 118–126.

Cline, H., & Haas, K. (2008). The regulation of dendritic arbor development and plasticity by glutamatergic synaptic input: A review of the synaptotrophic hypothesis. *Journal of Physiological, 586,* 1509–1517.

Cline, H. T., Wu, G. -Y., & Malinow, R. (1997). *In vivo* development of neuronal structure and function. *Cold Spring Harbor Symposia Quantitative Biology, LXI,* 95–104.

Da Silva, J. S., & Dotti, C. G. (2002). Breaking the neuronal sphere: Regulation of the actin cytoskeleton in neuritogenesis. *Nature Reviews. Neuroscience, 3,* 694–704.

Da Silva, J. S., Medina, M., Zuliani, C., Di Nardo, A., Witke, W., & Dotti, C. G. (2003). RhoA/ROCK regulation of neuritogenesis via profilin IIa-mediated control of actin stability. *The Journal of Cell Biology, 162,* 1267–1279.

Dacey, D. M. (1993). The mosaic of midget ganglion cells in the human retina. *The Journal of Neuroscience, 13,* 5334–5355.

Dailey, M. E., & Smith, S. J. (1996). The dynamics of dendritic structure in developing hippocampal slices. *The Journal of Neuroscience, 16,* 2983–2994.

Dalva, M. B., Takasu, M. A., Lin, M. Z., Shamah, S. M., Hu, L., Gale, N. W., & Greenberg, M. E. (2000). EphB receptors interact with NMDA receptors and regulate excitatory synapse formation. *Cell, 103,* 945–956.

Daub, H., Gevaert, K., Vandekerckhove, J., Sobel, A., & Hall, A. (2001). Rac/Cdc42 and p65PAK regulate the microtubule-destabilizing protein stathmin through phosphorylation at serine 16. *The Journal of Biological Chemistry, 276,* 1677–1680.

De Marco Garcia, N. V., Karayannis, T., & Fishell, G. (2011). Neuronal activity is required for the development of specific cortical interneuron subtypes. *Nature, 472,* 351–355.

Debant, A., Serra-Pages, C., Seipel, K., O'Brien, S., Tang, M., Park, S. H., & Streuli, M. (1996). The multidomain protein Trio binds the LAR transmembrane tyrosine phosphatase, contains a protein kinase domain, and has separate rac-specific and rho-specific guanine nucleotide exchange factor domains. *Proceedings of the National Academy of Sciences of the United States of America, 93,* 5466–5471.

Demyanenko, G. P., Halberstadt, A. I., Pryzwansky, K. B., Werner, C., Hofmann, F., & Maness, P. F. (2005). Abnormal neocortical development in mice lacking cGMP-dependent protein kinase I. *Brain Research. Development Brain Research, 160,* 1–8.

Devries, S. H., & Baylor, D. A. (1997). Mosaic arrangement of ganglion cell receptive fields in rabbit retina. *Journal of Neurophysiology, 78,* 2048–2060.

Dolmetsch, R. E., Pajvani, U., Fife, K., Spotts, J. M., & Greenberg, M. E. (2001). Signaling to the nucleus by an L-type calcium channel-calmodulin complex through the MAP kinase pathway. *Science, 294,* 333–339.

Eglen, S. J. (2006). Development of regular cellular spacing in the retina: Theoretical models. *Mathematical Medicine and Biology, 23,* 79–99.

Emoto, K., He, Y., Ye, B., Grueber, W. B., Adler, P. N., Jan, L. Y., & Jan, Y. N. (2004). Control of dendritic branching and tiling by the Tricornered-kinase/Furry signaling pathway in *Drosophila* sensory neurons. *Cell, 119,* 245–256.

Emoto, K., Parrish, J. Z., Jan, L. Y., & Jan, Y. N. (2006). The tumour suppressor Hippo acts with the NDR kinases in dendritic tiling and maintenance. *Nature, 443,* 210–213.

Etienne-Manneville, S. (2004). Actin and microtubules in cell motility: Which one is in control?. *Traffic, 5,* 470–477.

Famiglietti, E. V., Jr., & Kolb, H. (1976). Structural basis for ON-and OFF-center responses in retinal ganglion cells. *Science, 194,* 193–195.

Fuerst, P. G., Bruce, F., Tian, M., Wei, W., Elstrott, J., Feller, M. B., & Burgess, R. W. (2009). DSCAM and DSCAML1 function in self-avoidance in multiple cell types in the developing mouse retina. *Neuron, 64,* 484–497.

Fuerst, P. G., Harris, B. S., Johnson, K. R., & Burgess, R. W. (2010). A novel null allele of mouse DSCAM survives to adulthood on an inbred C3H background with reduced phenotypic variability. *Genesis, 48,* 578–584.

Fuerst, P. G., Koizumi, A., Masland, R. H., & Burgess, R. W. (2008). Neurite arborization and mosaic spacing in the mouse retina require DSCAM. *Nature, 451,* 470–474.

Fukata, M., Nakagawa, M., & Kaibuchi, K. (2003). Roles of Rho-family GTPases in cell polarisation and directional migration. *Current Opinion in Cell Biology, 15,* 590–597.

Furrer, M. P., Kim, S., Wolf, B., & Chiba, A. (2003). Robo and Frazzled/DCC mediate dendritic guidance at the CNS midline. *Nature Neuroscience, 6,* 223–230.

Gallegos, M. E., & Bargmann, C. I. (2004). Mechanosensory neurite termination and tiling depend on SAX-2 and the SAX-1 kinase. *Neuron, 44,* 239–249.

Gao, F. B., Brenman, J. E., Jan, L. Y., & Jan, Y. N. (1999). Genes regulating dendritic outgrowth, branching, and routing in *Drosophila*. *Genes and Development, 13,* 2549–2561.

Gao, F. B., Kohwi, M., Brenman, J. E., Jan, L. Y., & Jan, Y. N. (2000). Control of dendritic field formation in *Drosophila*: The roles of flamingo and competition between homologous neurons. *Neuron, 28,* 91–101.

Godinho, L., Mumm, J. S., Williams, P. R., Schroeter, E. H., Koerber, A., Park, S. W., & Wong, R. O. (2005). Targeting of amacrine cell neurites to appropriate synaptic laminae in the developing zebrafish retina. *Development, 132,* 5069–5079.

Goldberg, J. L. (2004). Intrinsic neuronal regulation of axon and dendrite growth. *Current Opinion in Neurobiology, 14,* 551–557.

Gonzalez, C. L., Gharbawie, O. A., Whishaw, I. Q., & Kolb, B. (2005). Nicotine stimulates dendritic arborization in motor cortex and improves concurrent motor skill but impairs subsequent motor learning. *Synapse, 55,* 183–191.

Govek, E. E., Newey, S. E., Akerman, C. J., Cross, J. R., Van der Veken, L., & Van Aelst, L. (2004). The X-linked mental retardation protein oligophrenin-1 is required for dendritic spine morphogenesis. *Nature Neuroscience, 7,* 364–372.

Govek, E. E., Newey, S. E., & Van Aelst, L. (2005). The role of the Rho GTPases in neuronal development. *Genes and Development, 19,* 1–49.

Grueber, W. B., Jan, L. Y., & Jan, Y. N. (2002). Tiling of the *Drosophila* epidermis by multidendritic sensory neurons. *Development, 129,* 2867–2878.

Grueber, W. B., Jan, L. Y., & Jan, Y. N. (2003). Different levels of the homeodomain protein cut regulate distinct dendrite branching patterns of *Drosophila* multidendritic neurons. *Cell, 112,* 805–818.

Grueber, W. B., & Sagasti, A. (2010). Self-avoidance and tiling: Mechanisms of dendrite and axon spacing. *Cold Spring Harbor Perspectives in Biology, 2,* a001750.

Grueber, W. B., Ye, B., Moore, A. W., Jan, L. Y., & Jan, Y. N. (2003b). Dendrites of distinct classes of *Drosophila* sensory neurons show different capacities for homotypic repulsion. *Current Biology, 13,* 618–626.

Gundersen, G. G., Gomes, E. R., & Wen, Y. (2004). Cortical control of microtubule stability and polarization. *Current Opinion in Cell Biology, 16,* 106–112.

Haas, K., Li, J., & Cline, H. T. (2006). AMPA receptors regulate experience-dependent dendritic arbor growth *in vivo*. *Proceedings of the National Academy of Sciences of the United States of America, 103,* 12127–12131.

Hall, A. C., Lucas, F. R., & Salinas, P. C. (2000). Axonal remodeling and synaptic differentiation in the cerebellum is regulated by WNT-7a signaling. *Cell, 100,* 525–535.

Hayashi, K., Ohshima, T., & Mikoshiba, K. (2002). Pak1 is involved in dendrite initiation as a downstream effector of Rac1 in cortical neurons. *Molecular and Cellular Neurosciences, 20,* 579–594.

Horton, A. C., & Ehlers, M. D. (2003). Dual modes of endoplasmic reticulum-to-Golgi transport in dendrites revealed by live-cell imaging. *The Journal of Neuroscience, 23,* 6188–6199.

Horton, A. C., & Ehlers, M. D. (2004). Secretory trafficking in neuronal dendrites. *Nature Cell Biology, 6,* 585–591.

Hossain, S., Hewapathirane, D. S., & Haas, K. (2012). Dynamic morphometrics reveals contributions of dendritic growth cones and filopodia to dendritogenesis in the intact and awake embryonic brain. *Developmental Neurobiology*, (in press).

Hua, J. Y., & Smith, S. J. (2004). Neural activity and the dynamics of central nervous system development. *Nature Neuroscience, 7,* 327–332.

Huang, C. S., Shi, S. H., Ule, J., Ruggiu, M., Barker, L. A., Darnell, R. B., & Jan, L. Y. (2005). Common molecular pathways mediate long-term potentiation of synaptic excitation and slow synaptic inhibition. *Cell, 123,* 105–118.

Huber, A. B., Kolodkin, A. L., Ginty, D. D., & Cloutier, J. F. (2003). Signaling at the growth cone: Ligand-receptor complexes and the control of axon growth and guidance. *Annual Review of Neuroscience, 26,* 509–563.

Hughes, M. E., Bortnick, R., Tsubouchi, A., Baumer, P., Kondo, M., Uemura, T., & Schmucker, D. (2007). Homophilic Dscam interactions control complex dendrite morphogenesis. *Neuron, 54,* 417–427.

Hume, R. I., & Purves, D. (1981). Geometry of neonatal neurones and the regulation of synapse elimination. *Nature, 293,* 469–471.

Hummel, T., Vasconcelos, M. L., Clemens, J. C., Fishilevich, Y., Vosshall, L. B., & Zipursky, S. L. (2003). Axonal targeting of olfactory receptor neurons in *Drosophila* is controlled by Dscam. *Neuron, 37,* 221–231.

Ikegaya, Y., Kim, J. A., Baba, M., Iwatsubo, T., Nishiyama, N., & Matsuki, N. (2001). Rapid and reversible changes in dendrite morphology and synaptic efficacy following NMDA receptor activation: Implication for a cellular defense against excitotoxicity. *Journal of Cell Science, 114,* 4083–4093.

Jan, Y. N., & Jan, L. Y. (2010). Branching out: Mechanisms of dendritic arborization. *Nature Reviews. Neuroscience, 11,* 316–328.

Jaworski, J., Spangler, S., Seeburg, D. P., Hoogenraad, C. C., & Sheng, M. (2005). Control of dendritic arborization by the phosphoinositide-3′-kinase-Akt-mammalian target of rapamycin pathway. *The Journal of Neuroscience, 25,* 11300–11312.

Jinushi-Nakao, S., Arvind, R., Amikura, R., Kinameri, E., Liu, A. W., & Moore, A. W. (2007). Knot/Collier and cut control different aspects

of dendrite cytoskeleton and synergize to define final arbor shape. *Neuron, 56,* 963–978.

Kay, J. N., Chu, M. W., & Sanes, J. R. (2012). MEGF10 and MEGF11 mediate homotypic interactions required for mosaic spacing of retinal neurons. *Nature, 483*, 465–469.

Kim, M. D., Jan, L. Y., & Jan, Y. N. (2006). The bHLH-PAS protein Spineless is necessary for the diversification of dendrite morphology of *Drosophila* dendritic arborization neurons. *Genes and Development, 20,* 2806–2819.

Kolb, B., Gorny, G., Li, Y., Samaha, A. N., & Robinson, T. E. (2003). Amphetamine or cocaine limits the ability of later experience to promote structural plasticity in the neocortex and nucleus accumbens. *Proceedings of the National Academy of Sciences of the United States of America, 100,* 10523–10528.

Kornhauser, J. M., Cowan, C. W., Shaywitz, A. J., Dolmetsch, R. E., Griffith, E. C., Hu, L. S., & Greenberg, M. E. (2002). CREB transcriptional activity in neurons is regulated by multiple, calcium-specific phosphorylation events. *Neuron, 34,* 221–233.

Krylova, O., Herreros, J., Cleverley, K. E., Ehler, E., Henriquez, J. P., Hughes, S. M., & Salinas, P. C. (2002). WNT-3, expressed by motoneurons, regulates terminal arborization of neurotrophin-3-responsive spinal sensory neurons. *Neuron, 35,* 1043–1056.

Kuntziger, T., Gavet, O., Manceau, V., Sobel, A., & Bornens, M. (2001). Stathmin/Op18 phosphorylation is regulated by microtubule assembly. *Molecular Biology of the Cell, 12,* 437–448.

Lamprecht, R., Farb, C. R., & LeDoux, J. E. (2002). Fear memory formation involves p190 RhoGAP and ROCK proteins through a GRB2-mediated complex. *Neuron, 36,* 727–738.

Le Roux, P., Behar, S., Higgins, D., & Charette, M. (1999). OP-1 enhances dendritic growth from cerebral cortical neurons *in vitro*. *Experimental Neurology, 160,* 151–163.

Lee, L. J., Lo, F. S., & Erzurumlu, R. S. (2005). NMDA receptor-dependent regulation of axonal and dendritic branching. *The Journal of Neuroscience, 25,* 2304–2311.

Lee, T., Winter, C., Marticke, S. S., Lee, A., & Luo, L. (2000). Essential roles of *Drosophila* RhoA in the regulation of neuroblast proliferation and dendritic but not axonal morphogenesis. *Neuron, 25,* 307–316.

Lefebvre, J. N., & Sanes, J. R. (2011). Gamma-protocadherins pattern starburst amacrine dendrites by self-avoidance. *Society for Neuroscience Abstracts*, S–7264.

Leslie, J. H., & Nedivi, E. (2011). Activity-regulated genes as mediators of neural circuit plasticity. *Progress in Neurobiology, 94,* 223–237.

Li, Z., Aizenman, C. D., & Cline, H. T. (2002). Regulation of rho GTPases by crosstalk and neuronal activity *in vivo*. *Neuron, 33,* 741–750.

Li, J., & Cline, H. T. (2010). Visual deprivation increases accumulation of dense core vesicles in developing optic tectal synapses in Xenopus laevis. *The Journal of Comparative Neurology, 518,* 2365–2381.

Li, J., Erisir, A., & Cline, H. (2011). *In vivo* time-lapse imaging and serial section electron microscopy reveal developmental synaptic rearrangements. *Neuron, 69,* 273–286.

Li, Z., Van Aelst, L., & Cline, H. T. (2000). Rho GTPases regulate distinct aspects of dendritic arbor growth in Xenopus central neurons *in vivo*. *Nature Neuroscience, 3,* 217–225.

Lisman, J., Schulman, H., & Cline, H. (2002). The molecular basis of CaMKII function in synaptic and behavioural memory. *Nature Reviews. Neuroscience, 3,* 175–190.

Luo, L. (2000). Rho GTPases in neuronal morphogenesis. *Nature Reviews. Neuroscience, 1,* 173–180.

Luo, L. (2002). Actin cytoskeleton regulation in neuronal morphogenesis and structural plasticity. *Annual Review of Cell and Developmental Biology, 18,* 601–635.

Matsuoka, R. L., Chivatakarn, O., Badea, T. C., Samuels, I. S., Cahill, H., Katayama, K., & Kolodkin, A. L. (2011a). Class 5 transmembrane semaphorins control selective Mammalian retinal lamination and function. *Neuron, 71,* 460–473.

Matsuoka, R. L., Nguyen-Ba-Charvet, K. T., Parray, A., Badea, T. C., Chedotal, A., & Kolodkin, A. L. (2011b). Transmembrane semaphorin signalling controls laminar stratification in the mammalian retina. *Nature, 470,* 259–263.

Matthews, B. J., Kim, M. E., Flanagan, J. J., Hattori, D., Clemens, J. C., Zipursky, S. L., & Grueber, W. B. (2007). Dendrite self-avoidance is controlled by Dscam. *Cell, 129,* 593–604.

Mauceri, D., Freitag, H. E., Oliveira, A. M., Bengtson, C. P., & Bading, H. (2011). Nuclear Calcium-VEGFD Signaling controls maintenance of dendrite arborization necessary for memory formation. *Neuron, 71,* 117–130.

McAllister, A. K., Lo, D. C., & Katz, L. C. (1995). Neurotrophins regulate dendritic growth in developing visual cortex. *Neuron, 15,* 791–803.

Montague, P. R., & Friedlander, M. J. (1991). Morphogenesis and territorial coverage by isolated mammalian retinal ganglion cells. *The Journal of Neuroscience, 11,* 1440–1457.

Morita, A., Yamashita, N., Sasaki, Y., Uchida, Y., Nakajima, O., Nakamura, F., & Goshima, Y. (2006). Regulation of dendritic branching and spine maturation by semaphorin3A-Fyn signaling. *The Journal of Neuroscience, 26,* 2971–2980.

Morrow, B. A., Elsworth, J. D., & Roth, R. H. (2005). Prenatal exposure to cocaine selectively disrupts the development of parvalbumin containing local circuit neurons in the medial prefrontal cortex of the rat. *Synapse, 56,* 1–11.

Mumm, J. S., Williams, P. R., Godinho, L., Koerber, A., Pittman, A. J., Roeser, T., & Wong, R. O. (2006). *In vivo* imaging reveals dendritic targeting of laminated afferents by zebrafish retinal ganglion cells. *Neuron, 52,* 609–621.

Na, E. S., & Monteggia, L. M. (2011). The role of MeCP2 in CNS development and function. *Hormones and Behavior, 59,* 364–368.

Nakayama, A. Y., Harms, M. B., & Luo, L. (2000). Small GTPases Rac and Rho in the maintenance of dendritic spines and branches in hippocampal pyramidal neurons. *The Journal of Neuroscience, 20,* 5329–5338.

Nedivi, E. (1999). Molecular analysis of developmental plasticity in neocortex. *Journal of Neurobiology, 41,* 135–147.

Nedivi, E., Fieldust, S., Theill, L. E., & Hevron, D. (1996). A set of genes expressed in response to light in the adult cerebral cortex and regulated during development. *Proceedings of the National Academy of Sciences of the United States of America, 93,* 2048–2053.

Nedivi, E., Hevroni, D., Naot, D., Israeli, D., & Citri, Y. (1993). Numerous candidate plasticity-related genes revealed by differential cDNA cloning. *Nature, 363,* 718–722.

Nedivi, E., Wu, G. Y., & Cline, H. T. (1998). Promotion of dendritic growth by CPG15, an activity-induced signaling molecule. *Science, 281,* 1863–1866.

Negishi, M., & Katoh, H. (2005). Rho family GTPases and dendrite plasticity. *Neuroscientist, 11,* 187–191.

Neves, G., Zucker, J., Daly, M., & Chess, A. (2004). Stochastic yet biased expression of multiple Dscam splice variants by individual cells. *Nature Genetics, 36,* 240–246.

Newey, S. E., Velamoor, V., Govek, E. E., & Van Aelst, L. (2005). Rho GTPases, dendritic structure, and mental retardation. *Journal of Neurobiology, 64,* 58–74.

Ng, J., Nardine, T., Harms, M., Tzu, J., Goldstein, A., Sun, Y., & Luo, L. (2002). Rac GTPases control axon growth, guidance and branching. *Nature, 416,* 442–447.

Niblock, M. M., Brunso-Bechtold, J. K., & Riddle, D. R. (2000). Insulin-like growth factor I stimulates dendritic growth in primary somatosensory cortex. *The Journal of Neuroscience, 20,* 4165–4176.

Nishimura, M., Gu, X., & Swann, J. W. (2011). Seizures in early life suppress hippocampal dendrite growth while impairing spatial learning. *Neurobiology of Disease, 44,* 205–214.

Nishiyama, M., Togashi, K., von Schimmelmann, M. J., Lim, C. S., Maeda, S., Yamashita, N., & Hong, K. (2011). Semaphorin 3A induces CaV2.3

channel-dependent conversion of axons to dendrites. *Nature Cell Biology, 13,* 676–685.

Nobes, C. D., & Hall, A. (1995). Rho, rac, and cdc42 GTPases regulate the assembly of multimolecular focal complexes associated with actin stress fibers, lamellipodia, and filopodia. *Cell, 81,* 53–62.

Novelli, E., Resta, V., & Galli-Resta, L. (2005). Mechanisms controlling the formation of retinal mosaics. *Progress in Brain Research, 147,* 141–153.

Palazzo, A. F., Cook, T. A., Alberts, A. S., & Gundersen, G. G. (2001). mDia mediates Rho-regulated formation and orientation of stable microtubules. *Nature Cell Biology, 3,* 723–729.

Palmer, A., & Klein, R. (2003). Multiple roles of ephrins in morphogenesis, neuronal networking, and brain function. *Genes and Development, 17,* 1429–1450.

Parrish, J. Z., Emoto, K., Jan, L. Y., & Jan, Y. N. (2007a). Polycomb genes interact with the tumor suppressor genes hippo and warts in the maintenance of *Drosophila* sensory neuron dendrites. *Genes and Development, 21,* 956–972.

Parrish, J. Z., Emoto, K., Kim, M. D., & Jan, Y. N. (2007b). Mechanisms that regulate establishment, maintenance, and remodeling of dendritic fields. *Annual Review of Neuroscience, 30,* 399–423.

Peichl, L., & Wassle, H. (1979). Size, scatter and coverage of ganglion cell receptive field centres in the cat retina. *Journal of Physiological, 291,* 117–141.

Penzes, P., Beeser, A., Chernoff, J., Schiller, M. R., Eipper, B. A., Mains, R. E., & Huganir, R. L. (2003). Rapid induction of dendritic spine morphogenesis by trans-synaptic ephrinB-EphB receptor activation of the Rho-GEF kalirin. *Neuron, 37,* 263–274.

Penzes, P., Johnson, R. C., Sattler, R., Zhang, X., Huganir, R. L., Kambampati, V., & Eipper, B. A. (2001). The neuronal Rho-GEF Kalirin-7 interacts with PDZ domain-containing proteins and regulates dendritic morphogenesis. *Neuron, 29,* 229–242.

Pfeffer, S., & Aivazian, D. (2004). Targeting Rab GTPases to distinct membrane compartments. *Nature Reviews. Molecular Cell Biology, 5,* 886–896.

Pilpel, Y., & Segal, M. (2004). Activation of PKC induces rapid morphological plasticity in dendrites of hippocampal neurons via Rac and Rho-dependent mechanisms. *The European Journal of Neuroscience, 19,* 3151–3164.

Polleux, F., & Ghosh, A. (2008). Molecular determinants of dendrite and spine development. In *Dentrites*, G. Stuart, N. Spruston, & M. Hausser (Eds.). Oxford, New York: Oxford University Press.

Polleux, F., Giger, R. J., Ginty, D. D., Kolodkin, A. L., & Ghosh, A. (1998). Patterning of cortical efferent projections by semaphorin-neuropilin interactions. *Science, 282,* 1904–1906.

Polleux, F., Morrow, T., & Ghosh, A. (2000). Semaphorin 3A is a chemoattractant for cortical apical dendrites. *Nature, 404,* 567–573.

Pratt, K. G., Dong, W., & Aizenman, C. D. (2008). Development and spike timing-dependent plasticity of recurrent excitation in the Xenopus optic tectum. *Nature Neuroscience, 11,* 467–475.

Purpura, D. P. (1974). Dendritic spine "dysgenesis" and mental retardation. *Science, 186,* 1126–1128.

Putz, U., Harwell, C., & Nedivi, E. (2005). Soluble CPG15 expressed during early development rescues cortical progenitors from apoptosis. *Nature Neuroscience, 8,* 322–331.

Rajan, I., & Cline, H. T. (1998). Glutamate receptor activity is required for normal development of tectal cell dendrites *in vivo*. *The Journal of Neuroscience, 18,* 7836–7846.

Rakic, P., Bourgeois, J. P., Eckenhoff, M. F., Zecevic, N., & Goldman-Rakic, P. S. (1986). Concurrent overproduction of synapses in diverse regions of the primate cerebral cortex. *Science, 232,* 232–235.

Redmond, L., Oh, S. R., Hicks, C., Weinmaster, G., & Ghosh, A. (2000). Nuclear Notch1 signaling and the regulation of dendritic development. *Nature Neuroscience, 3,* 30–40.

Rockhill, R. L., Euler, T., & Masland, R. H. (2000). Spatial order within but not between types of retinal neurons. *Proceedings of the National Academy of Sciences of the United States of America, 97,* 2303–2307.

Roska, B., & Werblin, F. (2001). Vertical interactions across ten parallel, stacked representations in the mammalian retina. *Nature, 410,* 583–587.

Rossman, K. L., & Sondek, J. (2005). Larger than Dbl: New structural insights into RhoA activation. *Trends in Biochemical Sciences, 30,* 163–165.

Rosso, S. B., Sussman, D., Wynshaw-Boris, A., & Salinas, P. C. (2005). Wnt signaling through Dishevelled, Rac and JNK regulates dendritic development. *Nature Neuroscience, 8,* 34–42.

Ruchhoeft, M. L., Ohnuma, S., McNeill, L., Holt, C. E., & Harris, W. A. (1999). The neuronal architecture of Xenopus retinal ganglion cells is sculpted by rho-family GTPases *in vivo*. *The Journal of Neuroscience, 19,* 8454–8463.

Ruthazer, E. S., Akerman, C. J., & Cline, H. T. (2003). Control of axon branch dynamics by correlated activity *in vivo*. *Science, 301,* 66–70.

Sanchez, A. L., Matthews, B. J., Meynard, M. M., Hu, B., Javed, S., & Cohen Cory, S. (2006). BDNF increases synapse density in dendrites of developing tectal neurons *in vivo*. *Development, 133,* 2477–2486.

Sanes, J. R., & Zipursky, S. L. (2010). Design principles of insect and vertebrate visual systems. *Neuron, 66,* 15–36.

Sasaki, Y., Cheng, C., Uchida, Y., Nakajima, O., Ohshima, T., Yagi, T., & Goshima, Y. (2002). Fyn and Cdk5 mediate semaphorin-3A signaling, which is involved in regulation of dendrite orientation in cerebral cortex. *Neuron, 35,* 907–920.

Schmidt, A., & Hall, A. (2002). Guanine nucleotide exchange factors for Rho GTPases: Turning on the switch. *Genes and Development, 16,* 1587–1609.

Schmucker, D., Clemens, J. C., Shu, H., Worby, C. A., Xiao, J., Muda, M., & Zipursky, S. L. (2000). *Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell, 101,* 671–684.

Schwartz, N., Schohl, A., & Ruthazer, E. S. (2009). Neural activity regulates synaptic properties and dendritic structure *in vivo* through calcineurin/NFAT signaling. *Neuron, 62,* 655–669.

Schwartz, N., Schohl, A., & Ruthazer, E. S. (2011). Activity-dependent transcription of BDNF enhances visual acuity during development. *Neuron, 70,* 455–467.

Scott, E. K., Reuter, J. E., & Luo, L. (2003). Small GTPase Cdc42 is required for multiple aspects of dendritic morphogenesis. *The Journal of Neuroscience, 23,* 3118–3123.

Shen, W., Da Silva, J. S., He, H., & Cline, H. T. (2009). Type A GABA-receptor-dependent synaptic transmission sculpts dendritic arbor structure in Xenopus tadpoles *in vivo*. *The Journal of Neuroscience, 29,* 5032–5043.

Shen, W., McKeown, C. R., Demas, J. A., & Cline, H. T. (2011). Inhibition to excitation ratio regulates visual system responses and behavior *in vivo*. *Journal of Neurophysiology, 106,* 2285–2302.

Shi, S. H., Cheng, T., Jan, L. Y., & Jan, Y. N. (2004a). The immunoglobulin family member dendrite arborization and synapse maturation 1 (Dasm1) controls excitatory synapse maturation. *Proceedings of the National Academy of Sciences of the United States of America, 101,* 13346–13351.

Shi, S. H., Cox, D. N., Wang, D., Jan, L. Y., & Jan, Y. N. (2004). Control of dendrite arborization by an Ig family member, dendrite arborization and synapse maturation 1 (Dasm1). *Proceedings of the National Academy of Sciences of the United States of America, 101,* 13341–13345.

Shima, Y., Kengaku, M., Hirano, T., Takeichi, M., & Uemura, T. (2004). Regulation of dendritic maintenance and growth by a mammalian 7-pass transmembrane cadherin. *Developmental Cell, 7,* 205–216.

Sin, W. C., Haas, K., Ruthazer, E. S., & Cline, H. T. (2002). Dendrite growth increased by visual activity requires NMDA receptor and Rho GTPases. *Nature, 419,* 475–480.

Soba, P., Zhu, S., Emoto, K., Younger, S., Yang, S. J., Yu, H. H., & Jan, Y. N. (2007). *Drosophila* sensory neurons require Dscam for dendritic self-avoidance and proper dendritic field organization. *Neuron, 54,* 403–416.

Sorensen, S. A., & Rubel, E. W. (2006). The level and integrity of synaptic input regulates dendrite structure. *The Journal of Neuroscience, 26,* 1539–1550.

Takasu, M. A., Dalva, M. B., Zigmond, R. E., & Greenberg, M. E. (2002). Modulation of NMDA receptor-dependent calcium influx and gene expression through EphB receptors. *Science, 295,* 491–495.

Threadgill, R., Bobb, K., & Ghosh, A. (1997). Regulation of dendritic growth and remodeling by Rho, Rac, and Cdc42. *Neuron, 19,* 625–634.

Tian, N., & Copenhagen, D. R. (2003). Visual stimulation is required for refinement of ON and OFF pathways in postnatal retina. *Neuron, 39,* 85–96.

Tolias, K. F., Bikoff, J. B., Burette, A., Paradis, S., Harrar, D., Tavazoie, S., & Greenberg, M. E. (2005). The Rac1-GEF Tiam1 couples the NMDA receptor to the activity-dependent development of dendritic arbors and spines. *Neuron, 45,* 525–538.

Van Aelst, L., & Cline, H. T. (2004). Rho GTPases and activity-dependent dendrite development. *Current Opinion in Neurobiology, 14,* 297–304.

Vaughn, J. E. (1989). Fine structure of synaptogenesis in the vertebrate central nervous system. *Synapse, 3,* 255–285.

Vaughn, J., Barber, R., & Sims, T. (1988). Dendritic development and preferential growth into synaptogenic fields: A quantitative study of Golgi-impregnated spinal motor neurons. *Synapse, 2,* 69–78.

Wang, J., Zugates, C. T., Liang, I. H., Lee, C. H., & Lee, T. (2002). *Drosophila* Dscam is required for divergent segregation of sister branches and suppresses ectopic bifurcation of axons. *Neuron, 33,* 559–571.

Wassle, H., Peichl, L., & Boycott, B. B. (1981). Dendritic territories of cat retinal ganglion cells. *Nature, 292,* 344–345.

Wassle, H., Puller, C., Muller, F., & Haverkamp, S. (2009). Cone contacts, mosaics, and territories of bipolar cells in the mouse retina. *The Journal of Neuroscience, 29,* 106–117.

Wassle, H., & Riemann, H. J. (1978). The mosaic of nerve cells in the mammalian retina. *Proceedings of the Royal Society of London. Series B, Biological Sciences, 200,* 441–461.

Wells, D. G., Dong, X., Quinlan, E. M., Huang, Y. S., Bear, M. F., Richter, J. D., & Fallon, J. R. (2001). A role for the cytoplasmic polyadenylation element in NMDA receptor-regulated mRNA translation in neurons. *The Journal of Neuroscience, 21,* 9541–9548.

Whitford, K. L., Marillat, V., Stein, E., Goodman, C. S., Tessier-Lavigne, M., Chedotal, A., & Ghosh, A. (2002). Regulation of cortical dendrite development by Slit-Robo interactions. *Neuron, 33,* 47–61.

Wittmann, T., & Waterman-Storer, C. M. (2001). Cell motility: Can Rho GTPases and microtubules point the way? *Journal of Cell Science, 114,* 3795–3803.

Wojtowicz, W. M., Flanagan, J. J., Millard, S. S., Zipursky, S. L., & Clemens, J. C. (2004). Alternative splicing of *Drosophila* Dscam generates axon guidance receptors that exhibit isoform-specific homophilic binding. *Cell, 118,* 619–633.

Wong, W. T., Faulkner-Jones, B. E., Sanes, J. R., & Wong, R. O. (2000). Rapid dendritic remodeling in the developing retina: Dependence on neurotransmission and reciprocal regulation by Rac and Rho. *The Journal of Neuroscience, 20,* 5024–5036.

Wu, G. Y., & Cline, H. T. (1998). Stabilization of dendritic arbor structure *in vivo* by CaMKII. *Science, 279,* 222–226.

Wu, G. Y., Deisseroth, K., & Tsien, R. W. (2001). Spaced stimuli stabilize MAPK pathway activation and its effects on dendritic morphology. *Nature Neuroscience, 4,* 151–158.

Wu, G., Malinow, R., & Cline, H. T. (1996). Maturation of a central glutamatergic synapse. *Science, 274,* 972–976.

Yacoubian, T. A., & Lo, D. C. (2000). Truncated and full-length TrkB receptors regulate distinct modes of dendritic growth. *Nature Neuroscience, 3,* 342–349.

Yamagata, M., & Sanes, J. R. (2008). Dscam and Sidekick proteins direct lamina-specific synaptic connections in vertebrate retina. *Nature, 451,* 465–469.

Yamagata, M., & Sanes, J. R. (2010). Synaptic localization and function of Sidekick recognition molecules require MAGI scaffolding proteins. *The Journal of Neuroscience, 30,* 3579–3588.

Yamagata, M., & Sanes, J. R. (2011). The immunoglobulin superfamily code for retinal laminar specificity: Genetic analysis of Sidekick-1 and 2 in mice. *Society for Neuroscience abstracts*, S–2158.

Yamagata, M., Weiner, J. A., & Sanes, J. R. (2002). Sidekicks: Synaptic adhesion molecules that promote lamina-specific connectivity in the retina. *Cell, 110,* 649–660.

Yu, T. W., & Bargmann, C. I. (2001). Dynamic regulation of axon guidance. *Nature Neuroscience, 4*(Suppl.), 1169–1176.

Yu, X., & Malenka, R. C. (2003). Beta-catenin is critical for dendritic morphogenesis. *Nature Neuroscience, 6,* 1169–1177.

Zheng, Y. (2004). G protein control of microtubule assembly. *Annual Review of Cell and Developmental Biology, 20,* 867–894.

Zhu, H., Hummel, T., Clemens, J. C., Berdnik, D., Zipursky, S. L., & Luo, L. (2006). Dendritic patterning by Dscam and synaptic partner matching in the *Drosophila* antennal lobe. *Nature Neuroscience, 9,* 349–355.

Zhu, J. J., Qin, Y., Zhao, M., Van Aelst, L., & Malinow, R. (2002). Ras and Rap control AMPA receptor trafficking during synaptic plasticity. *Cell, 110,* 443–455.

Zipursky, S. L., & Sanes, J. R. (2010). Chemoaffinity revisited: Dscams, protocadherins, and neural circuit assembly. *Cell, 143,* 343–353.

Hollis Cline, Anirvan Ghosh, Yuh-Nung Jan, and Josh Sanes

CHAPTER

21

Early Experience and Sensitive Periods

The nervous system has evolved to cope with an environment that is in many ways largely predictable. Therefore, much of the architecture and functional properties of the brain can be specified by genetic determinants that reflect the common experience of previous generations. Most of this circuitry is established prenatally, guided by genetically determined molecular mechanisms and shaped by patterns of spontaneous impulse activity that propagate through the central nervous system (CNS) in unborn animals, as discussed in Chapters 16, 18, and 19.

Not all aspects of an animal's world are certain, however. Details of an animal's physical characteristics vary, as do habitats and social conditions. To deal with such uncertainties, the CNS maintains the capacity to modify its connections based on the interactions of an animal with its environment. Through adaptive adjustments based on use or quality of performance, the developing nervous system customizes its functional properties to the needs and environment of the individual animal with a precision that does not need to be, and sometimes cannot be, encoded in the genome.

Although the nervous system is capable of making adaptive adjustments throughout the lifetime of an animal, many neural circuits pass through a period during their development when the capacity for adjustment in response to experience is substantially greater than it is after the circuit has matured. This period is referred to as a *sensitive period* (Knudsen, 2004). During a sensitive period, information derived from experience selects particular functional properties from a range of possible properties that a circuit could adopt. If appropriate experience is not gained during a sensitive period, many circuits never attain the ability to process information in a typical fashion, and, as a result, perception or behavior may be impaired permanently.

This chapter introduces four examples of circuits that have been relatively well studied with respect to their dependence on instruction by early experience for normal development. These examples are the circuits involved in (1) song memorization in songbirds, (2) sound localization in owls, (3) binocular processing in the mammalian visual cortex, and (4) temperament in rats. Sensitive periods for language learning in humans and filial imprinting in birds are also discussed. These examples are used to illustrate the principles that govern sensitive periods, discussed at the end of the chapter.

The terms *sensitive period* and *critical period* both appear in the scientific literature to denote a period of dramatically increased plasticity during development. *Sensitive period* is the more general term. The term *critical period* was borrowed (Hubel & Wiesel, 1970) from the behavioral literature, where it emphasizes that a particular kind of experience is essential (critical) for the development of a specific capacity or behavior, such as the capacity for stereoscopic vision or for language (Box 21.1). For many neural circuits, however, experience during this especially plastic period may not be essential for the circuit to develop the capacity to function adaptively. The latter example would be a sensitive but not a critical period. Thus, while all critical periods are sensitive periods, not all sensitive periods are critical.

As you read about each of these examples, you will notice a number of recurring themes. First, as circuits that are shaped by experience during a sensitive period develop, they reach a stage when they rely on ("expect") specific spatial and temporal patterns of input in order to select a particular pattern of interconnections. These learned patterns of interconnections adjust the properties of the circuit to the sensory conditions of the individual animal. Second, the changes in the pattern of interconnections that result last for the lifetime of the animal. Third, sensitive period learning begins when the circuit is capable of processing the relevant information. This capacity requires that both excitatory and inhibitory circuitry

© 2013 Elsevier Inc. All rights reserved.

BOX 21.1

SENSITIVE PERIODS IN HUMANS

Many human capabilities depend critically on experience gained during early life. These capabilities range from fundamental capacities, such as stereoscopic vision, visual acuity, and binocular coordination, to high-level capacities, such as social behavior, language, and the ability to perceive forms and faces. Because in each case, specific kinds of experience during a restricted period in early life is essential for the normal development of the capacity, these periods are sometimes referred to as critical periods.

The best known and most thoroughly studied sensitive period in humans is for language (see Chapter 48). A clear relationship exists between the age of exposure to a language and the level of proficiency achieved in that language (Newport, Bavelier, & Neville, 2001). This relationship holds for the learning of both first and second languages. Acquisition of a first language has been assessed in children who have been raised in the absence of any language (feral or abused children) or, more frequently, in congenitally deaf children who have been raised without the aid of sign language. Much more data are available for people who began learning a second language at different ages. For both first and second languages, a thorough command of the language is attained by those who learn the language before 7 years of age. The degree of language proficiency that is eventually achieved decreases progressively with age of exposure and reaches adult levels by the end of adolescence.

Only certain aspects of language are affected by learning during sensitive periods (Kuhl, 2000; Newport et al., 2001). Full proficiency with grammar (the classes of words, their functions, and relations in a sentence), syntax (the way in which words are put together in a sentence), and the production and comprehension of phonetics (the speech sounds of a language) are each dependent on early exposure to language. In contrast, semantics (word meaning) and size of vocabulary are not affected by the age of exposure. Thus, sensitive periods seem to affect the formal and subtle aspects of language, whereas the capacity to learn new words and their meanings continues unabated throughout life.

Physiological measures reveal an age dependence in the way in which language is processed and represented in the brain (Dehaena et al., 1997; Weber-Fox & Neville, 1996). Various techniques have been used to assess brain activity while human subjects make perceptual judgments in language tasks. These techniques include functional magnetic resonance imaging, positron emission tomography, and event-related potentials. In normal adults, language is processed in specific areas, primarily in the left or "dominant" hemisphere. In people who have learned a second language prior to the age of 7 years, the brain areas that are involved in processing the first and second languages overlap extensively. In contrast, in people who have learned a second language later in life, the areas of the brain that are activated by the second language do not overlap, or they overlap little, with those that are activated by the first language. The brain areas activated by the second language are less lateralized to the left hemisphere and are more variable across subjects. The effect of age at the time of learning on the brain areas activated by language is far more conspicuous for tasks requiring grammatical and phonic judgments than for tasks requiring semantic judgments. Thus, consistent with the behavioral observations of the age dependence for learning grammar and phonetics, the regions of the brain that contribute to the processing of grammar and phonetics are shaped in a unique way during sensitive periods.

Detailed knowledge of the mechanisms that control sensitive periods and of the plasticity that occurs during sensitive periods will provide a basis for formulating optimal therapeutic procedures to help minimize long-term harmful effects of early abnormal experience—associated with neonatal and childhood disabilities, for example—and maximize the acquisition of normal function once normal conditions are restored. Such knowledge may also lead to improved methods of rearing and teaching normal children that take advantage of the full capacity of the central nervous system to learn from experience.

Eric I. Knudsen

References

Dehaena, S., Doupoux, E., Mehler, J., Cohen, L., Perani, D., van de Moortele, P. -F., … Le Bihan, D. (1997). Anatomical variability in the cortical representation of first and second languages. *Neuroreport, 17,* 3809–3815.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences of the United States of America, 97,* 11850–11857.

Newport, E. L., Bavelier, D., & Neville, H. J. (2001). Critical thinking about critical periods: Perspectives on a critical period for language acquisition. In E. Doupoux (Ed.), *Language, brain and cognitive development: Essays in honor of Jacques Mehler* (pp. 481–502). Cambridge, MA: MIT Press.

Weber-Fox, C. M., & Neville, H. J. (1996). Maturational constraints on functional specializations for luanguage processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience, 8,* 231–256.

operate effectively and that the molecular mechanisms necessary to support synaptic remodeling are active. Fourth, sensitive period learning ends once an animal has received adequate experience and the neural circuit has become committed to a particular pattern of interconnections. Finally, the timing of sensitive periods can be altered by environmental conditions that deprive the animal of the necessary input or that arouse the animal in the context of the relevant experience or by manipulation of the cellular or molecular machinery that governs the plasticity. These examples of sensitive period learning highlight the dynamic interplay between genetic instruction and the profound influence of early experience in shaping the functional capacities of the mature brain.

BIRDSONG: LEARNED BY EXPERIENCE

The song of most songbirds depends on learning that occurs early in life during a sensitive period (Doupe & Kuhl, 1999; Konishi, 1985). Birdsong is a special form of vocal communication used by certain species of birds to identify neighbors, defend territories, and attract mates. Songs are distinguished from other communication sounds by their length, spectral complexity, and periodic structure—properties that give birdsong its melodic quality. The songs sung by birds are characteristic of the species (conspecific song); dialects of the species' song often reflect the geographical area in which the bird was raised (Fig. 21.1A).

Songs are passed on from one generation to the next by a combination of genetic instruction and learning. In a few species of songbirds, the influence of genetic instruction is strong, and learning plays a relatively minor role. In most, however, the role of learning is paramount. Some of these species learn new songs each year (seasonal learners), whereas others learn their songs only once early in life. In the latter case, the birds memorize their songs as juveniles, during a sensitive period.

White-crowned sparrows and zebra finches are species that memorize their songs during a sensitive period (Immelmann, 1972; Konishi, 1985). The extent of the sensitive period in each species has been determined by raising birds in acoustic isolation and then exposing them to conspecific song for brief periods in development. The effect of this experience on song learning is assessed by observing the song that the male eventually sings (in these species, only the male sings). Birds raised in acoustic isolation throughout the sensitive period sing an "isolate" song (Fig. 21.1B) that lacks the spectral and temporal complexity typical of normal song. When juvenile birds are allowed to hear conspecific song even for a few days during the sensitive period, however, they memorize that particular song and reproduce it accurately when they later learn to sing.

Song memorization in these species illustrates an important principle that pertains to most sensitive period learning: The nervous system is genetically predisposed to accept only a limited range of potential stimuli as appropriate for learning (Doupe & Kuhl, 1999). For example, juvenile birds that are allowed to hear only alien songs that differ substantially from their conspecific song develop isolate song, indicating that they reject these distinctly alien songs as models for learning. Even within the range of songs that a bird will learn, it strongly prefers a conspecific song when given a choice of several similar song types. Moreover, juveniles memorize conspecific songs rapidly, whereas they memorize slightly different alien songs only after much longer periods of experience. Thus, the circuits responsible for song memorization contain genetically determined filters that require certain spectral and temporal features before the stimulus is accepted as

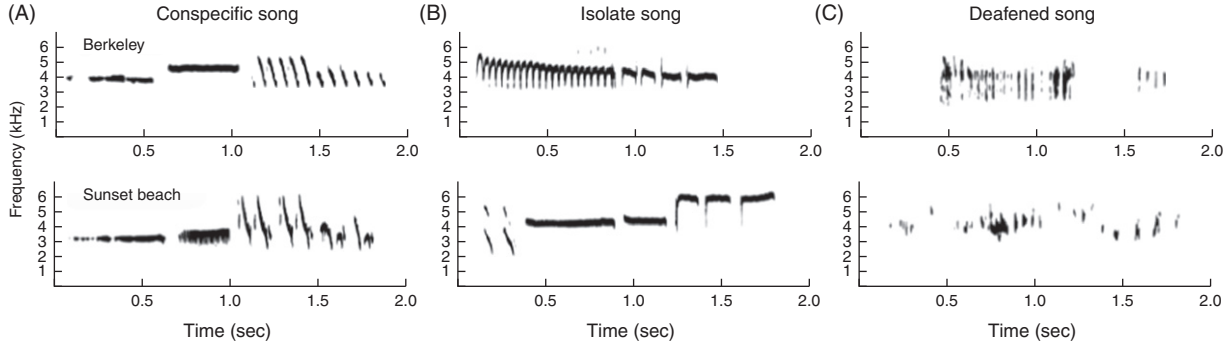


FIGURE 21.1 Songs of white-crowned sparrows. These are sonograms (time-frequency sound spectrograms) of songs from birds with different kinds of early experience. Sound energy in each frequency band is indicated by the darkness of the trace. (A) Song dialects. Birds raised in different areas sing slightly different songs. These dialects are stable for many years and are transmitted by learning. (B) Isolate songs. These simpler songs develop in birds raised in acoustic isolation or in birds that fail to copy a tutor song. (C) Songs of deafened birds. These kinds of songs develop in birds that are deafened after the sensitive period for song memorization but before the period of vocal learning. The birds need to hear their own voice to develop normal song. *From Konishi (1985).*

appropriate, and within the range of stimuli that is deemed acceptable, some song patterns are preferred over others.

Song learning involves two components: song memorization and vocal learning. In white-crowned sparrows, these components are separated by many months (Fig. 21.2A), whereas in zebra finches, which develop much more rapidly, they overlap. During the sensitive period for song memorization, according to a current hypothesis, high-order sensory neurons become tuned to respond selectively to the acoustic patterns of the songs that the bird memorizes. During the period of vocal learning, these high-order neurons act as templates for evaluating the bird's own song, guiding the development of song so that it eventually matches the previously memorized song pattern.

A Sensitive Period for Song Memorization

The sensitive period for song memorization begins at about 2 weeks of age and lasts for about 8 weeks in both zebra finches and white-crowned sparrows (Doupe & Kuhl, 1999; Immelmann, 1972). Baby birds that are exposed to a conspecific song before the sensitive period opens do not learn the song, even though they can hear at this early age. This suggests that the neuronal substrate for song memorization is not yet ready to be shaped by experience. In addition, juvenile birds that do not hear a normal song until after 3 to 4 months of age do not learn to sing a normal song. Apparently, by this age, the influence of experience on the neuronal substrate for song memorization has become reduced greatly.

The time at which learning occurs during the sensitive period for song memorization is determined by the individual's own experience. Once the sensitive period has opened, exposure of a juvenile bird to normal song for 1 week is sufficient for the bird to memorize the song, although subsequent exposure to other songs can still modify the song that the bird comes to sing. If a juvenile bird is kept in acoustic isolation (or hears only songs that are suboptimal as models for learning) for many weeks past the opening of the sensitive period and then hears normal song, it learns the normal song. Thus, during the sensitive period, the nervous system waits in a receptive state for appropriate experience-dependent instruction. Once this instruction is received, a particular pattern of connectivity becomes established and the sensitive period closes (Knudsen, 2004).

When a juvenile is deprived continuously of appropriate auditory experience, its capacity to memorize song eventually diminishes with age. Under these conditions, the sensitive period closes gradually because of additional, age-dependent factors (discussed later) that reduce the plasticity of the relevant circuits. As a bird approaches this age, experience with appropriate stimuli must be richer in order to have an effect. For example, white-crowned sparrows raised in acoustic isolation until 50 days of age no longer memorize songs presented from loudspeakers, but do memorize songs presented by live tutors. Thus, the enriched sensory experience provided by social interactions with the tutor overcomes the decline in the facility of the pathway for song memorization, essentially extending the sensitive period.

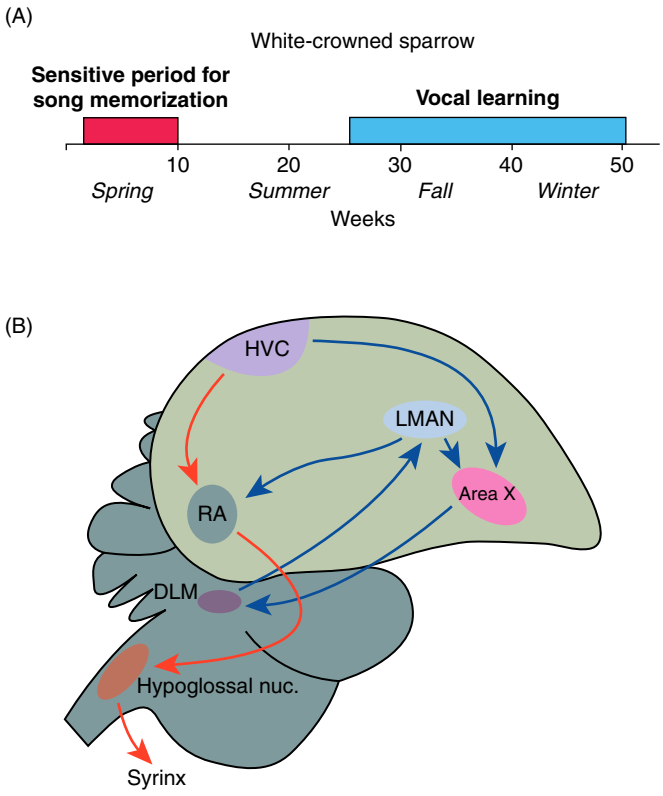


FIGURE 21.2 The sensitive period for song memorization and the set of nuclei that comprise the song system in songbirds. (A) Time-line for song learning for the white-crowned sparrow. For this species, the sensitive period for song memorization does not overlap with the period when the bird learns to sing. (B) A schematic diagram of a side view of the brain of a songbird. The vocal motor pathway, shown in red, consists of the HVC, the robust nucleus of the archopallium (RA), and the hypoglossal nucleus. The anterior basal ganglia pathway, shown in blue, consists of Area X, the medial portion of the dorsolateral nucleus of the thalamus (DLM), and the lateral portion of the magnocellular nucleus of the anterior nidopallium (LMAN).

Vocal Learning

Learning to sing requires a combination of vocal practice and auditory feedback. A young bird that is deafened after song memorization but before the onset of vocal learning, and is thereby prevented from hearing its own voice, will not develop a normal song (Fig. 21.1C). Clearly, auditory feedback is essential for shaping the patterns of connectivity in the vocal motor pathway while the bird is learning to sing. Auditory feedback is necessary for the

bird both to learn how internally generated motor commands correspond with the sounds they produce and to compare the sounds they produce with the memorized song template.

A Neural Pathway for Song Learning

The neural mechanisms that underlie song memorization are being explored in many laboratories but, as yet, we know little about this aspect of song learning. In contrast, a great deal is known about the neural mechanisms that underlie song production (Mooney, 2009). The pathway for song production was identified by its sexual dimorphism in species in which only males sing. In these species, a distinct set of nuclei, referred to as the song system, is conspicuously hypertrophied in males (Fig. 21.2B). The song system consists of two distinct groups of nuclei: one group in the posterior forebrain that is responsible for song production and another in the anterior forebrain that is critical for song learning and maintenance.

The posterior, vocal motor pathway consists of three serially connected nuclei: the HVC and the robust nucleus of the archopallium (RA) in the forebrain and the hypoglossal nucleus in the brainstem (Fig. 21.2B). As a bird prepares to sing, a wave of neural activity spreads from the HVC to the RA, and finally to the hypoglossal nucleus, which contains the motor neurons that control the vocal musculature. Bilateral lesions of the HVC or the RA leave birds permanently incapable of producing song, although they still can make other kinds of unlearned vocalizations.

The anterior, basal ganglia pathway (Fig. 21.2B) consists of Area X, a thalamic nucleus (DLM), and the lateral portion of the magnocellular nucleus of the anterior nidopallium (LMAN), and is essential for experience-dependent adjustments of song (Brainard & Doupe, 2000; Sober & Brainard, 2009). Lesions made in the LMAN of a young bird that is just learning to sing cause a dramatic cessation of song development, freezing the bird's song in an immature state. This freezing of song resembles song crystallization. In adult birds, LMAN lesions result in birds that sing normal songs, but can no longer adjust the quality of their songs based on experience.

In young birds, the anterior pathway provides information to the vocal motor pathway for the purpose of vocal learning. According to one hypothesis, the anterior pathway compares auditory feedback about the song that a bird produces with a stored template of the memorized song. The result of this comparison is then used to instruct the development and maintenance of connections in the vocal motor pathway. Consistent with this hypothesis, Area X and the LMAN in the anterior pathway contain neurons that respond maximally to the sound of the bird's own song. During development, axons from the LMAN are the first to innervate the RA in the vocal motor pathway. Information transmitted by the LMAN-RA pathway is mediated predominantly by NMDA receptors, a class of glutamate receptors that, when activated, is known in other systems to induce synaptic plasticity. As birds begin learning to sing, a second set of axons enters the RA from the HVC in the vocal motor pathway and begins making glutamatergic synapses. These later connections may be guided by the activity of the preexisting, LMAN-RA synapses.

Hormonal Regulation of Learning

A bird's song becomes stable ("crystallizes") by the time it reaches adulthood. Indeed, exposure of a juvenile bird to high levels of testosterone causes its song to crystallize prematurely in an abnormal state (Konishi, 1985). Conversely, juvenile birds that are castrated before they learn to sing produce inconsistent song patterns throughout life. In adult birds, consistent song is maintained by the action of the anterior pathway, which detects and corrects deviations of the song that the bird is singing from the memorized song (Sober & Brainard, 2009).

Seasonal song learners, such as canaries, appear to recapitulate the process of song memorization and vocal learning each year. This relearning is linked with, and could result from, the waxing and waning of steroid hormone levels. This raises the intriguing possibility that these sensitive periods can be opened and closed by hormonal or environmental factors.

Summary

Song learning in birds shares many characteristics with language learning in humans (Box 21.1). Song is a learned form of vocal communication used by certain species of birds to attract mates, defend their territories, and identify their neighbors. Songs are learned in two phases. The first is song memorization, which for many species takes place during a sensitive period. The second is vocal learning, during which vocal practice and auditory feedback shape the bird's song to match the memorized song. The neural pathway for song production is sexually dimorphic in many species in which only males sing. A separate pathway, in the anterior forebrain, plays a special role in vocal learning, guiding adjustments in sound production as the bird learns to sing and actively maintaining the quality of adult song over time.

SOUND LOCALIZATION: CALIBRATED BY EARLY EXPERIENCE IN THE OWL

A pathway that is highly modifiable during a sensitive period is the auditory pathway that creates a map of space in the midbrain of the barn owl (Fig. 21.3). This pathway

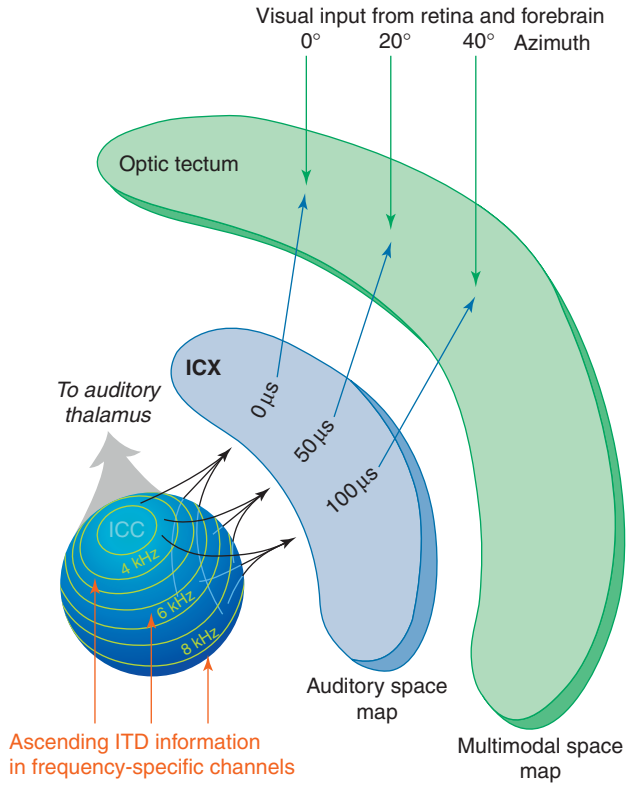


FIGURE 21.3 The ascending auditory pathway to the optic tectum in the barn owl. Auditory inputs enter the optic tectum from the brainstem (bottom arrows); these inputs already encode frequency-specific information about interaural time difference (ITD). These inputs project into the central nucleus of the inferior colliculus (ICC), where they are organized topographically by frequency. Bands in the ICC represent these frequencies (e.g., 4, 6, 8 kHz). Neurons in the ICC convey information both to the auditory thalamus (the primary pathway) and to the external nucleus of the inferior colliculus (ICX). In the ICX, ITD information is combined across frequency channels to synthesize a map of auditory space. For example, an ITD of 0 µs is generated by sound stimuli directly in front of the animal, such that sound reaches the two ears at exactly the same time. At the position marked 0 µs in the ICX are neurons that respond maximally to sounds with an ITD value of 0 µs and thus respond selectively to sounds originating in front of the animal. Sounds originating, for example, from positions that are further to the left-hand side will reach the ears with progressively greater left ear-leading ITDs and thus stimulate neurons with progressively larger best ITDs. From the ICX, the auditory map of space is conveyed via a topographic projection to the optic tectum. Here the auditory map is aligned and merged with a visual map of space (top arrows, representing inputs from the retina and the forebrain) to produce a multimodal space map.

transforms a representation of auditory spatial cues that exists in the central nucleus of the inferior colliculus (ICC; see Chapter 24) into a topographic representation of space in the external nucleus of the inferior colliculus (ICX). The auditory map of space is then sent on to the optic tectum, the avian analog of the mammalian superior colliculus, where it aligns with and is integrated with a visual map of space. The function of this pathway is to determine the locations of auditory stimuli, information that can be used to direct the animal's attention and orienting movements toward interesting auditory stimuli (see Chapter 31).

The pathway derives the location of a sound source by evaluating spatial cues that are present in the auditory signals at the two ears. The most reliable cues for sound localization are interaural timing differences (ITDs) and interaural level differences (ILDs). ITDs arise from the difference in the path length that sound must travel to reach the near versus the far ear. Because the ears are on the sides of the head, ITD varies systematically with the horizontal (azimuthal) location of a sound source (Fig. 21.4A; contour lines). ILDs result from the fact that each ear is most sensitive to sound coming from certain directions. Consequently, a sound from a particular direction will usually produce a higher sound level in one ear than the other. ILDs vary both with the azimuthal and with the elevational location of a sound source in spatial patterns that depend on sound frequency.

In creating the map of auditory space, the nervous system can only roughly anticipate the relationship between encoded values of ITD and ILD and the locations of sound sources that produce them. The correspondence of ITDs and ILDs with locations in space changes with the size and shape of the head and ears, features that vary across individuals, as well as for a given individual during growth. Moreover, the encoded values of sound timing and level that are transmitted to the CNS depend on the sensitivity and transduction properties of each ear, and these properties can change over time. Therefore, to establish and maintain an accurate map of space in the optic tectum, this midbrain pathway must learn the exact relationship between the encoded cue values and the locations of sound sources that produce them.

The influence of early experience on the owl's auditory space map has been demonstrated using a variety of techniques that change the relationship between cue values and locations in space: For example, the external ears have been altered drastically or the auditory canal of one ear has been plugged chronically (Knudsen, 1999). The midbrain pathway responds adaptively to such manipulations by adjusting the tuning of neurons in the ICX and optic tectum to ITDs and ILDs that restore an accurate map of space. In young animals, this plasticity enables the recovery of a substantially normal auditory map even after severe disruptions of hearing. In adult animals, plasticity is far more limited in extent.

An instructive signal that adjusts the tuning of ICX and tectal neurons is provided by the visual system. The instructive role of vision in guiding the tuning of ICX and tectal neurons has been demonstrated in experiments in which owls wear optical displacing prisms that chronically shift the visual field (Knudsen, 2002). The effect of experience with displacing prisms on auditory spatial tuning is most apparent in the optic tectum, where the visual receptive field of each neuron indicates the location

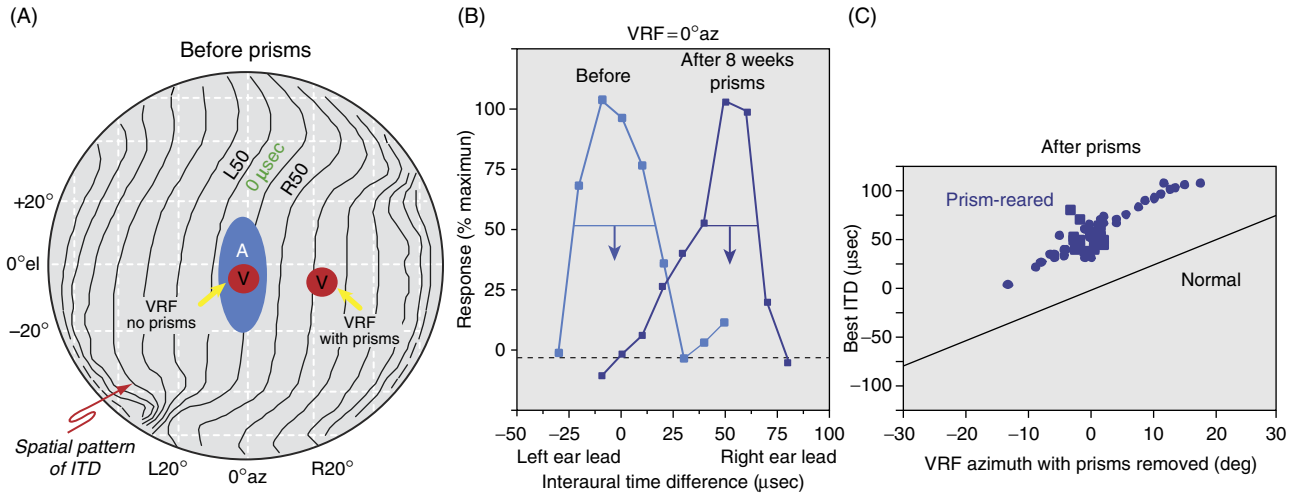


FIGURE 21.4 Rearing owls with laterally displacing, optical prisms causes an adaptive shift in the tuning of neurons in the optic tectum for ITD. (A) This map represents the space in front of the owl, showing both the elevation and the azimuth of a stimulus in space. Contour lines indicate the correspondence of ITD values (in microseconds) with particular locations in space. The point at which the 0° axes intersect represents the point in space directly in front of the owl's head. The auditory (A) and visual (V) receptive fields of one tectal neuron are shown in the center of the map. This neuron responds optimally when the stimulus is directly in front of the animal. Normally, the auditory and visual receptive fields are aligned. Optical prisms induce a horizontal displacement of the neuron's visual receptive field (VRF), resulting in a misalignment between A and V. (B) Tuning for ITD is shifted by prism experience. These ITD tuning curves were recorded from similar sites in the optic tectum before (blue) and after (purple) 8 weeks of prism experience. Both sites had a VRF at 0° azimuth. After 8 weeks of experience, the neuron is tuned for the ITD produced by an acoustic stimulus at the location of the optically displaced VRF, as shown in A. Arrows indicate the best ITD for each site; the best ITD is defined as the center of the range of ITDs to which the neuron responded with more than 50% of its maximum response. (C) The relationship between best ITD and VRF azimuth is shifted systematically from normal in prism-reared owls. The black line indicates the regression of best ITD on VRF azimuth that is observed in normal owls. Dots represent individual sites in a prism-reared owl. The map of ITD is shifted systematically relative to the visual map of space.

in auditory space to which that neuron should normally be tuned. When prisms that displace the visual field horizontally are placed in front of the eyes, the visual receptive fields of tectal neurons are shifted horizontally and out of alignment with the auditory receptive fields of these neurons (Fig. 21.4A). In juvenile birds, continuous experience with such prisms over a period of 6 to 8 weeks causes the auditory receptive fields of tectal neurons to realign with their visual receptive fields: The tuning of tectal neurons to ITDs and ILDs changes so that they respond to auditory cue values representing the locations of their optically displaced visual receptive fields (Figs. 22.4B, C). As a result, the auditory map of space shifts to match the optically shifted visual map. This adjustment is adaptive because, by making it, the animal alters its orientation toward sounds so that it sees the source of the sound through the prisms.

A Sensitive Period for Neuronal Adjustments

This adaptive auditory plasticity is regulated developmentally. The magnitude of the shift in neuronal ITD tuning that is induced under standard conditions of prism experience depends greatly on the age of the animal (Fig. 21.5). Large shifts in ITD tuning, of up to 70 μs, occur in juvenile owls (even larger shifts can be induced by ear

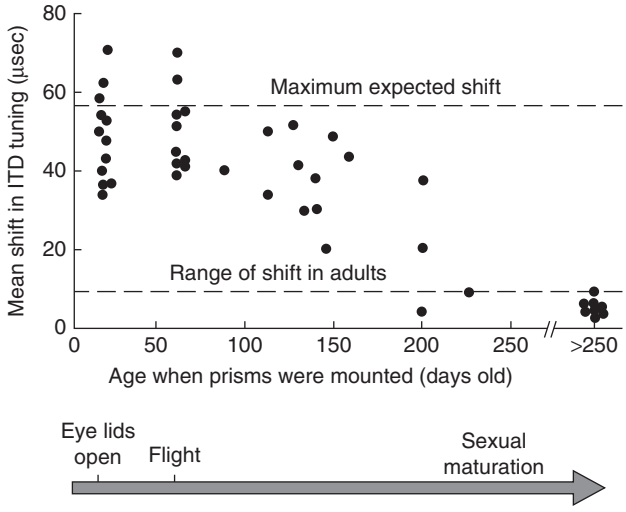


FIGURE 21.5 The sensitive period for visual calibration of neuronal ITD tuning in the optic tectum. Each dot represents data from a single owl. The large arrow below is a timeline, indicating important developmental stages in an owl's life. Each owl experienced a 23° displacement of the visual field for at least 60 days. ITD tuning was then measured at 15 to 23 sites in the superficial layers of the optic tectum. The difference between the best ITD measured and the best ITD expected normally, based on the location of the site's VRF (see Fig. 21.4C), was taken as the "shift in ITD tuning." The mean shift in ITD tuning for the population of sampled sites as a function of the age of the owl when prisms were first mounted is plotted.

plugging). In adult owls, equivalent conditions rarely shift ITD tuning by more than 10 μs, even after many months of prism experience. The period during which prism experience induces large changes in ITD tuning, the sensitive period, ends as the owls approach sexual maturity, at about 200 to 250 days old.

Although an experience-dependent shift in ITD tuning is observed most easily in the optic tectum (due to the physiological reference provided by the visual receptive field of the neuron), the site in the pathway where the plasticity actually takes place is in the ICX: The maps of ITD in the ICX and in the optic tectum are shifted by equivalent amounts in prism-reared owls, whereas the representation of ITD in the ICC remains unaltered.

Mechanisms of Plasticity

A shift of the auditory space map that occurs during the sensitive period is associated with a change in the architecture of the neurons that project from the ICC to the ICX (DeBello, Feldman, & Knudsen, 2001). A topographic projection from the ICC to the ICX brings ITD information to the appropriate site in the ICX, where the information is integrated across frequency channels to create spatial receptive fields (Fig. 21.3). A topographic projection is established early in development, before prism experience exerts its effects (Fig. 21.6A). Prism experience causes neurons in the portion of the ICC that represents the shifted values of ITD to project axons to neurons in novel regions of the ICX (Fig. 21.6B; red lines) that then become tuned to those abnormal values of ITD. Thus, experience induces the elaboration of axons at sites in the ICX where they support appropriate responses. In owls that have acquired shifted space maps, the learned anatomical circuit coexists with the normal circuit.

The newly learned responses that result from prism experience are mediated differentially by special classes of excitatory and inhibitory neurotransmitter receptors. Drugs that specifically block the *n*-methyl-d-aspartate (NMDA) glutamate receptor, such as AP5, eliminate or severely reduce the responses of ICX neurons to newly learned values of ITD, while having substantially less effect on their responses to the normal value. At the same time, responses to the now inappropriate, normal values of ITD are actively suppressed by powerful inhibition mediated by $GABA_A$ receptors. The actions of NMDA and $GABA_A$ receptors play critical roles in many other examples of experience-dependent plasticity as well.

As mentioned earlier, large shifts in ITD tuning in response to visual field displacement occur only in juvenile owls during a sensitive period. In contrast, removal of prisms from adult owls that have been raised from the day of eye opening wearing prisms results in a shift of the map of ITD back to normal. The genetically programmed, normal circuitry persists into adulthood even without validation by experience. In owls that have been raised with prisms, the adult circuit is able to switch back and forth, over a period of weeks, from an abnormal representation of ITD to a normal representation of ITD and vice versa, depending on the visual world the animal experiences. In this case, acquired alterations in circuit architecture resulting from experience during the sensitive period, together with the genetically programmed circuit architecture, influence the range of connectional states that the circuit can assume later in adult life.

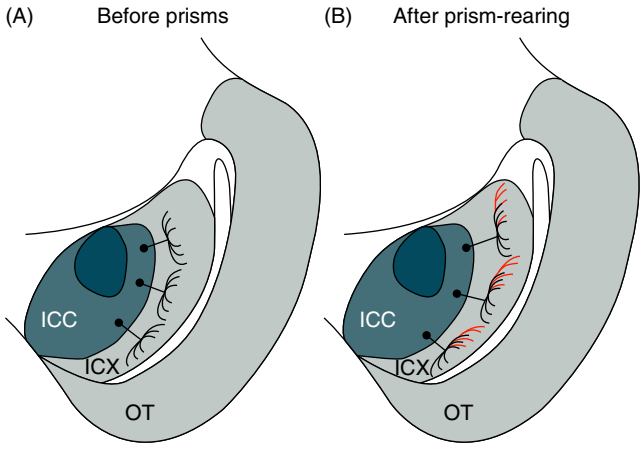


FIGURE 21.6 Schematic model of the change in the pattern of axonal projections from the ICC to the ICX that accompanies the shift in the map of ITD in the ICX. Based on DeBello et al. (2001). (A) The initial state of the projection before prism experience. (B) After prism experience, axonal projections from the ICC to the ICX are shifted systematically as indicated by the red arbors.

Plasticity after the Sensitive Period

Although plasticity is greatest during this sensitive period, the capacity for substantial plasticity persists into adulthood. The difference is that, in adult owls, more sophisticated conditions are necessary in order to induce plasticity. For example, relatively large cumulative changes in the functional properties of neurons can be instructed in small steps over time, as long as each incremental step is within the more limited adjustment range of the adult pathway. Also, plasticity can be increased dramatically in adults when the owls are highly attentive to the information that instructs change, such as when they hunt live prey that they can both see and hear. These special conditions are not necessary to induce large-scale plasticity during the sensitive period. Nevertheless, these conditions engage mechanisms of plasticity that drive adaptive adjustments in the adult nervous system, and they could be optimized to maximize adaptive changes in adult animals after sensitive periods have closed.

Summary

The brain derives the location of a sound source by evaluating spatial cues that are present in the auditory signals at the two ears. Interaural timing differences and interaural level differences are used to determine the position of a sound in space. In owls, this information is used to create a map of auditory space, which is aligned closely with a visual map of space in the optic tectum. Establishing and maintaining the alignment of the auditory space map with the visual space map in the optic tectum is an active process, guided by information provided by experience. The capacity of this pathway to change adaptively in response to experience is particularly great during a sensitive period. Plasticity during the sensitive period involves changes in circuit architecture and depends on the action of specific subtypes of glutamate and GABA receptors.

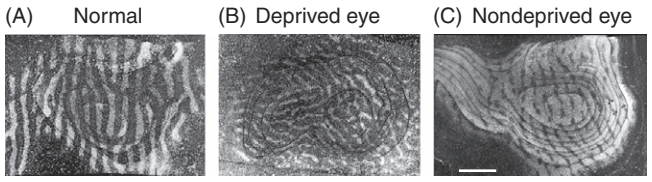


FIGURE 21.7 Critical period plasticity of the anatomical projections representing the left and right eyes in the visual cortex of monkeys. These are dark field autoradiographs of tangential sections through layer 4 of the primary visual cortex, showing the pattern of trans-synaptic labeling (bright areas) from the thalamic lateral geniculate nucleus (LGN) that resulted from an injection of a radioactive tracer into one eye; the intervening dark areas represent LGN input from the opposite (non-injected) eye. Scale bar indicates 2.5 mm. (A) The pattern of LGN input to the visual cortex in a normal monkey. (B) Severely reduced LGN input from the deprived eye of a monkey that experienced monocular deprivation from 3 weeks until 7 months old. (C) Greatly expanded LGN input from the nondeprived eye of a monkey that experienced monocular deprivation from 2 weeks until 18 months old. *From Hubel et al. (1977).*

BINOCULAR PROCESSING IN THE MAMMALIAN VISUAL CORTEX DEPENDS ON EARLY EXPERIENCE

In the mammalian nervous system, visual information from the two eyes first comes together at the level of the primary visual cortex. As described in detail in Chapter 18, visual experience during a limited period early in life has an enormous impact on how much of the visual cortex is devoted to processing input from each eye and the degree to which binocular inputs are combined (Hubel & Wiesel, 1970; Hubel, Wiesel, LeVay, 1977). Early in development (before birth in many species), afferents that provide inputs from the left and right eyes, respectively, begin to cluster in separate, interleaved areas in layer 4 of the primary visual cortex (Fig. 21.7A). This early clustering of eye-specific inputs from the lateral geniculate nucleus (LGN) is driven by a combination of molecular mechanisms and patterns of spontaneous neuronal activity.

Soon after birth, a period, referred to as a "critical period" by Hubel and Wiesel (1970), opens during which visual experience influences the competition among LGN afferents for territory in layer 4. As long as the eyes are coordinated and used equally, the typical final state, consisting of equally wide ocular dominance columns, is achieved (Fig. 21.7A). If, however, vision is impaired in one eye, due to monocular eyelid closure in an experimental animal or to a cataract in a human, for example, the balance between LGN afferents in their competition for territory in layer 4 is disrupted: LGN afferents that convey input from the impaired eye lose the ability to drive layer 4 neurons in an abnormally large region of the cortex (Fig. 21.7B), whereas LGN afferents that convey input from the nondeprived eye gain the ability to drive layer 4 neurons in an abnormally large portion of the cortex (Fig. 21.7C). As a consequence, activity throughout most of the visual cortex becomes driven by LGN afferents from the nondeprived eye.

A Critical Period for Binocular Processing in the Visual Cortex

The critical period for binocular processing in the visual cortex has been studied particularly carefully in cats and primates by measuring the age dependence of the effects of monocular eyelid closure. The onset of the critical period is rapid, beginning at about 3 weeks in cats. By 4 to 6 weeks of age, the cortex is maximally sensitive to monocular deprivation: A few days of monocular deprivation cause a complete shift in ocular dominance, leaving the cortex almost entirely driven by input from the nondeprived eye (Fig. 21.8). Beyond 6 weeks of age, the critical period gradually closes: Over the next 10 months, the rate at which ocular dominance can be shifted by monocular deprivation and the degree to which it can be shifted both decrease (Fig. 21.9). Once a cat is about 1 year old, monocular deprivation even for months no longer affects ocular dominance in the cortex (Fig. 21.8C). Similar critical periods, but extending later in life, exist for the visual cortex in monkeys and humans.

The critical period for binocular processing in the visual cortex is unusual in one respect: There is no predisposition to establish a normal pattern of connectivity based on normal experience. Even after baby cats or monkeys have experienced normal binocular vision for several weeks, monocular deprivation still causes the responses in the cortex to become dominated by the nondeprived eye. Moreover, once the responses in the cortex become

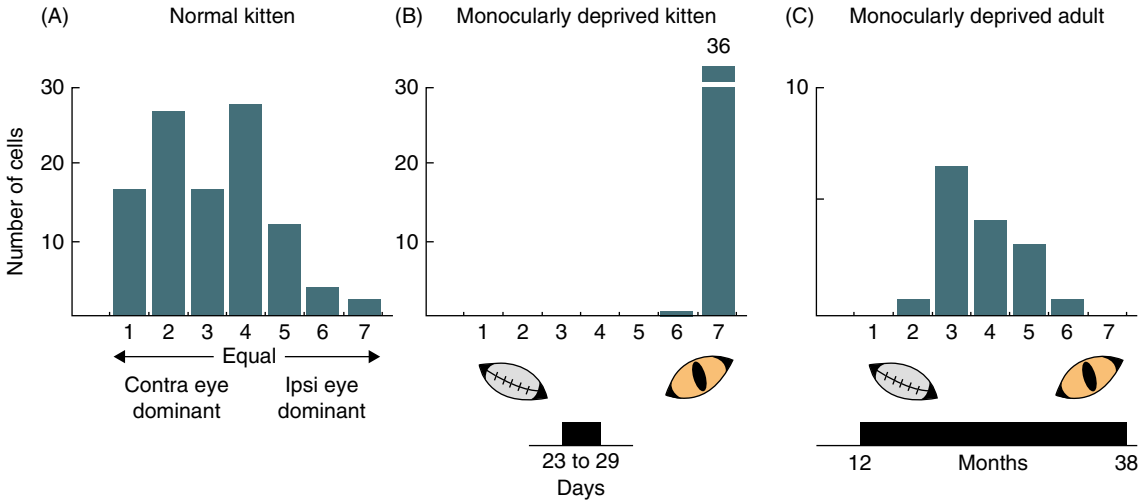


FIGURE 21.8 Effect of chronic closure of one eye on the responsiveness of visual cortical neurons to input from each eye. (A) Ocular dominance distribution in the primary visual cortex of two normal kittens, 3 to 4 weeks old. Cells in group 1 were driven only by the contralateral eye; for group 2, the contralateral eye was markedly dominant; for group 3, the contralateral eye was slightly dominant; for group 4, there was no apparent difference in the drive from the two eyes; for group 5, the ipsilateral eye dominated slightly; for group 6, it dominated markedly; and for group 7, cells were driven only by the ipsilateral eye. (B) Ocular dominance distribution was altered dramatically in a kitten exposed to contralateral eye closure for 1 week (from 23 to 29 days of age). (C) Ocular dominance distribution was essentially normal in an adult cat exposed to contralateral eye closure for 26 months. *From Hubel and Wiesel (1970).*

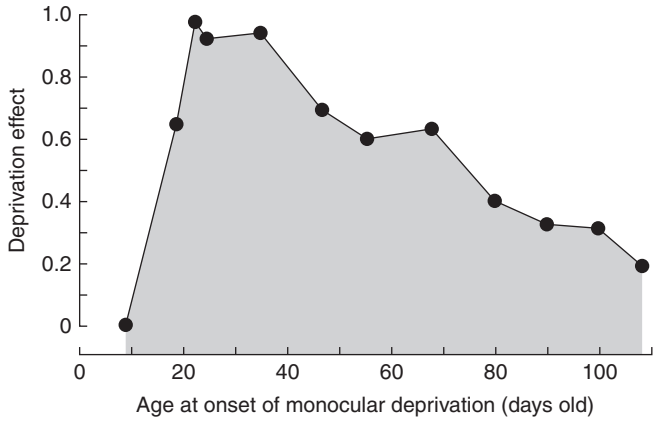


FIGURE 21.9 The critical period for binocular processing in the primary visual cortex of the cat. The degree of functional disconnection of cortical neurons from the deprived eye is quantified and plotted as a function of the kitten's age at the time of monocular closure. Chronic monocular closure lasted 10 to 12 days. Each point represents data from a single animal. Functional disconnection was based on the ocular dominance distribution (see Fig. 21.8) and indicated the degree to which the influence of the closed eye was weakened or lost. The index was defined such that the mean value for normal cats was 0, whereas total disconnection resulted in a value of 1. *From Olson and Freeman (1980).*

dominated by one eye, reinstating normal visual input to the previously deprived eye does not readily restore normal binocular responses in the cortex. Instead, the nondeprived eye continues to dominate the responses of cortical neurons for as long as animals have been studied (up to 5 years after restoration of binocular input). Thus, unlike song memorization in songbirds, where just a brief exposure to conspecific song establishes a robust and permanent memory of the song, the neural circuitry that supports normal binocular processing is not stabilized rapidly by exposure to normal binocular input. Instead, normal binocular vision must persist throughout the entire critical period to prevent the cortex from becoming dominated by monocular responses. An adaptive advantage of this characteristic of the binocular pathway has yet to be recognized.

Research on ocular dominance plasticity has shifted recently to rodents, particularly to mice, in which targeted genetic manipulations have led to major discoveries regarding molecular mechanisms that govern the plasticity (Tropea, Van Wart, Sur, 2009). However, the critical period is far less critical in rodents than it is in cats or primates (Hofer, Mrsic-Flogel, Bonjoeffer, & Hubener, 2006): Monocular deprivation causes significant shifts in ocular dominance in the cortex of adult mice, and restoration of binocular vision after monocular deprivation during the critical period results in the recovery of binocular responses. Thus, as with song memorization in songbirds, the degree to which binocular processing in the visual cortex depends on experience during a restricted developmental period varies across species.

Mechanisms of Plasticity

The dramatic change in the responsiveness of neurons in the visual cortex to input from the two eyes that is caused by early monocular deprivation is accompanied by an equally dramatic change in the axonal architecture

of the LGN neurons that project to layer 4: axonal arbors conveying input from the normal eye expand while those conveying input from the impaired eye shrink. The anatomical remodeling of LGN axons depends on the availability of neurotrophins—BDNF and NT-4—as well as on the expression of their cognate receptor, TrkB (Lein & Shatz, 2001). The adjustments of synaptic drive by the left and right eyes result from long-term potentiation and long-term depression of synapses in the cortex, and require the activation of NMDA receptors.

A variety of mechanisms that contribute to synapse formation or stabilization have been shown to modulate ocular dominance plasticity. Ocular dominance plasticity is enhanced by genetic overexpression of BDNF in mice or by application of neuromodulators, such as acetylcholine or norepinephrine, in cats. Plasticity can be enhanced in both young and adult mice by genetic mutations of developmentally regulated molecules that normally restrict plasticity, such as a classically immune system molecule and its receptor (PirB) or Nogo and its receptor, which inhibits axonal outgrowth, or by enzymatic degradation or genetic manipulation of the extracellular matrix, which restricts synaptic remodeling (Miyata, Komatsu, Yoshimura, Taya, & Kitagawa, 2012). Most of these interventions have been shown to prolong plasticity in developing animals or to "reopen" ocular dominance plasticity in adults.

The spatial and temporal patterns of activity that drive cortical plasticity depend on a precise balance of excitatory and inhibitory influences. The critical period for binocular processing does not open until inhibitory connections, which develop after excitatory connections, become effective (Hensch, 2005). The critical period opens prematurely in mice in which inhibition has been enhanced pharmacologically. Conversely, the critical period is delayed indefinitely in mice in which a class of GABA receptors has been genetically deleted and, in these same mice, ocular dominance plasticity is reinstated by pharmacological enhancement of GABA-mediated transmission.

Large, rapid shifts in ocular dominance in response to monocular deprivation occur only during a critical period. However, in adult mice, in which a shift in ocular dominance has been induced by monocular deprivation during the critical period and is followed by recovery of binocular responses by reopening the eye, closure of the same eye for a second time (but not of the nondeprived eye) causes a rapid shift in ocular dominance in the same direction (Hofer et al., 2006). Thus, analogous to the case described for barn owl sound localization, changes in circuit architecture that result from experience during the critical period have a permanent effect on the range of connectional states that the visual cortex can assume later in adult life.

Dark-Rearing Prolongs the Critical Period

An environmental manipulation that prolongs the critical period is raising animals in complete darkness. At the beginning of the critical period, most cortical neurons respond to inputs from either eye, but the responses tend to be weak. In cats that are raised in the dark until well past the end of the critical period (as defined by monocular deprivation), cortical neurons continue to be driven binocularly and their responses remain weak. When these animals are finally allowed visual experience, the responses of these neurons gradually increase in strength and the relative representations of the left and right eyes are shaped by the animal's experience: In monocularly deprived cats, nondeprived eye inputs become predominant in the cortex, and in cats that experience binocular vision, normal ocular dominance columns develop. Thus, without strong experience-driven input, the mechanisms that control ocular representation in the cortex remain in an uncommitted state, waiting for instruction for a prolonged, if not indefinite, period of time.

Summary

The processing of input from the two eyes in the primary visual cortex is shaped by binocular experience during a sensitive period. Because the effects of disruptions of binocular vision are apparently irreversible when they occur during this limited period in early life, this period has been referred to as a *critical period*. During prenatal development and before the onset of vision, thalamic inputs representing the left and right eyes segregate from each other to form ocular dominance columns of roughly equal width in the visual cortex. Soon after birth, a period opens during which the visual experience of the animal shapes the representations of the two eyes in the cortex. Normal binocular vision consolidates and refines the established patterns of ocular representation. Impaired vision in one eye, however, causes the LGN axons in layer 4 carrying information from the impaired eye to shrink and their synapses to loose efficacy, whereas axons carrying information from the nonimpaired eye expand and their synapses increase in efficacy. In contrast, equivalent, monocular deprivation in adult animals has no apparent effect, at least in cats and primates. The critical period for binocular processing in the visual cortex can be prolonged substantially by rearing animals in complete darkness.

A SENSITIVE PERIOD FOR SHAPING THE TEMPERAMENT OF RATS

The way in which a rat pup is cared for by its mother during a sensitive period has an enormous impact on a rat's temperament as an adult (Weaver et al., 2004).

In this example, the primary influence of early experience is to set the expression level of a gene, the gene for the glucocorticoid receptor (GR), in a particular circuit in the hippocampus. The protein product of this gene strongly influences a rat's emotional responses to stressful conditions. The effect of mothering on the expression level of this gene occurs only during the first week of a rat's life, and the behavioral consequences of this experience last throughout adulthood.

Different rats respond differently to threatening situations. For example, when introduced into a new environment, some rats are calm and adventurous, whereas others are anxious and fearful. This fundamental difference in emotional responses to stressors reflects, in part, a rat's interactions with its mother during the first week after birth (Fig. 21.10). Rats raised by a mother who groomed them extensively (high-grooming) and nursed them in a way that facilitated their access to milk (arched-back nursing) are less fearful and less reactive to stressors as adults than are rats raised by a mother who did not treat them in this way.

Cross-fostering experiments demonstrate that the transmission of these traits is dominated by experience and not by genetics. Rats born to low-grooming mothers (nonattentive and no arched-back nursing) but raised by high-grooming mothers become themselves calm, adventurous, high-grooming mothers. Conversely, rats born to high-grooming mothers but raised by low-grooming mothers become anxious, low-grooming mothers. Thus, the transmission of these traits depends on mother–infant interactions during the first week of a rat's life. The experiments demonstrate that, although genetics constrains the ranges of these traits, early experience can modify them dramatically.

FIGURE 21.10 A rat mother grooms her pups during the first week after their birth. This interaction permanently affects the temperament of these rat pups.

The emotional responses of a rat to stressors reflect the reactivity of the hypothalamic-pituitary-adrenal (HPA) system. Activation of the HPA system leads to the release of a special class of stress hormones, glucocorticoid hormones, from the adrenal glands. Animals with high circulating levels of glucocorticoid hormones are anxious and fearful, those with low circulating levels of this hormone are calm.

The hippocampus exerts a powerful negative feedback on the release of glucocorticoid hormones from the adrenal glands. High levels of GRs in the hippocampus result in low basal levels of glucocorticoids and tight regulation of glucocorticoid release when an animal is stressed. Low levels of GRs in the hippocampus have the opposite effects.

Mechanisms of Plasticity

The expression level of GRs in the hippocampus is adjusted by interactions of a rat pup with its mother. Rats raised by high-grooming mothers express high levels of GRs in the hippocampus, whereas rats raised by low-grooming mothers express low levels of GRs. The same effect is observed in experiments in which mothering behavior is manipulated. For example, removing rat pups from a mother for 15 minutes and then returning them to the mother, causes the mother to increase her grooming of those pups. Rats that experience this increase in grooming express increased levels of GRs in the hippocampus compared with control animals; correspondingly, they are less reactive to stressors, less fearful, and more adventurous. This manipulation has these effects only when applied during the first two weeks after birth.

One mechanism that mediates the effect of mothering on GR levels in the hippocampus, and therefore on an animal's reactivity to stressors, is the methylation of the GR gene. Methylation silences the GR gene by blocking access of transcription factors to a promoter region of the gene. Gene promoter methylation is adjusted during development and is stable thereafter. Experience during the first week after birth with a high-grooming mother (Fig. 21.10) causes the GR gene promoter to become demethylated, resulting in high levels of GR gene expression and high levels of GRs in the hippocampus. Conversely, experience with low-grooming mothers causes the GR gene promoter to become methylated, leading to low GR levels in the hippocampus. The sensitive period for the effect of mothering on the temperament of rats is accounted for by the narrow window in development when the methylation state of the GR gene is adjusted by experience. The persistence of the effects of this experience is due to the subsequent stability of the gene's methylation state.

A second mechanism that contributes to the effect of mothering on GR levels in the hippocampus is the

acetylation of histones. Acetylation of histones in the chromatin of a gene increases the access of transcription factors to promoters. Histones within the nucleosome core that includes the GR gene are more acetylated in rats reared by high-grooming mothers than in rats reared by low-grooming mothers, an effect that increases the transcription of the GR gene and raises GR levels in the hippocampus. The difference in histone acetylation is thought to be driven by the difference in the methylation state of the GR gene, discussed earlier.

The contribution of histone acetylation to controlling GR levels in the hippocampus has been tested by administering an inhibitor of histone deacetylation to adult rats that had been reared by low-grooming mothers. The inhibitor increased acetylation of the histones surrounding the GR gene, increasing access of demethylating enzymes to the gene. The inhibitor thereby caused a decrease in the methylation state of the gene, which increased GRs in the hippocampus and changed the temperament of the adult rats into the calm, adventurous temperament typical of rats raised by high-grooming mothers.

Summary

Interactions of a rat with its mother during the first week of life shape the way in which a rat responds to stressful situations as an adult. The primary mechanism for this effect is an experience-dependent adjustment of the expression level of the GR gene in the hippocampus. High levels of maternal care during the first postnatal week causes a demethylation of a GR gene promoter, resulting in a stable increase in GR gene expression in the hippocampus and a consequent decrease in basal levels of glucocorticoid hormones and a tighter regulation of glucocorticoid hormone release in response to stressors. As a result, the animal is calm, adventurous, and reacts in a measured fashion to stressors. Maternal behavior has these effects only during a sensitive period in a rat's development when the methylation state of the GR gene can be altered. Thereafter, the methylation state of the GR gene remains stable throughout adulthood.

PRINCIPLES OF SENSITIVE PERIOD LEARNING

During the maturation of many neural circuits, patterns of neuronal activity, driven by stimuli or the animal's behavior, shape the circuit's functional properties, architecture, and/or biochemistry. This shaping process selects circuit properties that are appropriate for the individual's experience.

Circuit changes that occur during sensitive periods differ from those that occur in adulthood in terms of their magnitude and persistence and the behavioral conditions under which the changes can occur. As illustrated by the preceding examples in this chapter, sensitive period experience can cause changes in the nervous system that are beyond the range of changes that occur in adults, and the changes that result typically persist for the lifetime of the animal.

Changes in circuit properties are more readily induced during sensitive periods than in mature circuits. During sensitive periods, circuits may be altered simply by exposing animals to unusual conditions such as monocular deprivation or abnormal patterns of auditory spatial cues, for example. In mature circuits, equivalent conditions either have no effect or require the attention of the animal to the conditions in order for plasticity to occur.

Sensitive period learning is influenced heavily by genetic predispositions. Only a limited range of stimuli is allowed to operate as an instructive influence for a particular circuit. Within this acceptable range, some stimuli are preferred over others. This property is well illustrated by song memorization in songbirds and by imprinting (Box 21.2). The predisposition of the nervous system to be instructed by "normal" experience probably originates in the selectivity of the response properties, genetically determined as well as shaped by experience, of the neurons that provide input to the sites in the pathway where the learning takes place. As learning progresses, the selectivity of the pathway for acceptable input increases.

Whether a particular pathway passes through a sensitive period can vary across species. For example, some songbirds, such as canaries and mockingbirds, learn new songs seasonally throughout life, whereas others, such as white-crowned sparrows and zebra finches, learn their songs only during a sensitive period. Such species differences may provide a useful tool for uncovering the mechanisms that are responsible for sensitive periods.

The magnitude of changes that may result from experience-driven adjustments varies greatly across circuits and across species. The magnitude of changes depends on the degree of genetic specification of the inputs to the site of change. When the selection of appropriate inputs is from a large potential range of inputs, the effect of experience can have a profound influence on a circuit. Conversely, when the range of potential inputs is highly restricted by genetic specification, the effect of experience is correspondingly small.

The duration of different sensitive periods also varies greatly. At one end of the spectrum are the sensitive periods for imprinting, which may open and close within hours (Box 21.2). At the other end are sensitive periods for acquiring complex cognitive capabilities, such as language (described in Chapter 48), which involve sensitive periods that last for many years.

BOX 21.2

FILIAL IMPRINTING: BABIES LEARN TO RECOGNIZE THEIR PARENTS

For many species of birds and mammals, including ducks, geese, mice, and monkeys, parental care is essential for the survival of the young. The young of these species learn rapidly to distinguish their parents from all other individuals and form a unique and close relationship with their parents from that point on—a process referred to as filial imprinting (Hess, 1973; Horn, 2004). Filial imprinting can involve the learning of visual, auditory, olfactory, and gustatory cues that identify a parent. The learning of these cues takes place during short, well-defined sensitive periods early in postnatal life.

The visual component of the learning process usually is preceded by auditory, olfactory, and/or gustatory components. In many species, the babies learn to recognize the vocalizations of the mother based on experience that begins before or soon after the animal is born. In addition, babies may learn the odor and/or taste of the mother from the odors and tastes experienced immediately after birth. The recognition of the parent based on acoustic and/or chemical cues helps the young select the correct individual for visual imprinting, once the eyes and nervous system are capable of adequate form vision.

Sensitive periods for filial imprinting are relatively discrete and can be as short as a few hours in some species (Ramsay & Hess, 1954). For example, the sensitive period for filial imprinting in ducklings occurs during the first day of life. When a duckling is exposed once, for 10 minutes, to a model of a male duck on the first day after hatching, the duckling imprints on this particular model. When the duckling is tested 5 to 70 hours later, by being offered a choice between the previously presented model and a model of a female duck, the duckling prefers to follow the model of the male duck over that of the female duck, even when the female model is much closer to the duckling and makes louder calls. Measured by the "following response," imprinting in ducks is most effective between 10 and 20 hours after hatching.

As in other examples of learning during sensitive periods, young animals exhibit an innate preference to imprint on normal stimuli. When given a choice, babies in the process of imprinting attend preferentially to images that more closely resemble members of their own species. Thus, when baby ducks are given the choice of imprinting on geese or on people, they imprint on the (duck-like) geese. This predisposition is based on genetically programmed preferences for simple, conspicuous features—referred to as sign stimuli by ethologists—that tend to distinguish the species from all others. This implies that the neural circuitry involved in filial imprinting, like that involved in song learning in songbirds, contains genetically determined neuronal filters that help identify stimuli that are appropriate models for learning. As imprinting proceeds, learning causes these filters to become more selective until ultimately the young are capable of discriminating one individual from all others.

Imprinting results in rapid functional and structural changes in a specific part of the forebrain of birds (Horn, 2004). In chicks that have been imprinted on an artificial object, neurons in the intermediate and medial mesopallium (IMM) of the forebrain become responsive to the imprinted stimulus, and some become highly selective for the stimulus. Associated with the acquisition of selective neuronal responses in the IMM are local increases in the number of NMDA receptors, increases in the size of certain synapses, and a temporary surge in inhibitory activity. The increased inhibitory activity, which begins soon after exposure to the imprinting stimulus and lasts for many hours, is thought to contribute critically to the shaping of the specificity of IMM neurons for the imprinted stimulus.

Eric I. Knudsen

References

Hess, E. H. (1973). *Imprinting: Early experience and the developmental psychobiology of attachment*. New York, NY: Van Nostrand-Reinhold.

Horn, G. (2004). Pathways of the past: The imprint of memory. *Nature reviews Neuroscience, 5*, 108–120.

Ramsay, A. O., & Hess, E. H. (1954). A laboratory approach to the study of imprinting. *Wilson Bulletin, 66*, 196–206.

The Opening of Sensitive Periods Depends on Pathway Maturation

Sensitive periods cannot open until the brain has matured to the point where three conditions are met. First, the information provided to the circuit from lower level circuits must be sufficiently reliable and precise to allow the circuit to carry out its function. Second, the circuit's connectivity must have matured adequately for it to process the information. For example, the critical period in the visual cortex does not open until and unless inhibitory as well as excitatory connections are effective. Third, the mechanisms that enable plasticity must be active.

Because complex behaviors depend on information that is processed through hierarchies of circuits, the first condition implies that sensitive periods for circuits at

higher levels in these hierarchies cannot open until the information from circuits at lower levels has become reliable. Sensitive periods for low-level circuits, such as the LGN, occur earlier than sensitive periods for higher level circuits, such as those in the visual cortex. The same principle holds for circuits at different levels in the owl's sound localization pathway. Thus, pathways that support complex behaviors, such as human language (Box 21.1) or object recognition, may rely on circuits that pass through sensitive periods that end at very different stages in an animal's life.

Experience cannot shape a circuit until the mechanisms that enable plasticity are active. Experience-driven changes in functional properties can involve a remodeling of axons and dendrites, an elaboration of new synapses, an adjustment of synaptic efficacies, and/or a regulation of gene expression. Anatomical remodeling and the establishment of new synapses may well be guided by the same mechanisms that control synaptogenesis during circuit development (Chapter 16). Neurotrophins (described in Chapter 17)—particularly BDNF, for example—are essential for anatomical remodeling during the critical period in the visual cortex (Lein & Shatz, 2001). Mechanisms that regulate synaptic efficacy during a sensitive period include those that underlie synaptic plasticity in the adult nervous system (Chapter 47). Long-term potentiation (LTP) and long-term depression (LTD), for example, have been shown to operate in many models of sensitive period learning. Moreover, the ubiquitous presence of NMDA receptors at sites of change in the various models of developmental learning indicates that this class of glutamate receptor plays a key role in adjusting patterns of connectivity during sensitive periods.

The Closing of Sensitive Periods Can Involve Many Mechanisms

Sensitive periods end once an animal has received adequate experience and the relevant circuit is irreversibly committed to a pattern of connectivity. The factors that render the commitment irreversible are not known. Many factors may play a role, and the factors that are most important may differ across different pathways.

One factor that may contribute to the closing of sensitive periods is the age- or experience-dependent alteration of molecular mechanisms that support changes in synaptic efficacy such as LTP and LTD, described in Chapter 46. In addition, for those circuits in which axonal elaboration is required for experience-dependent changes in connectivity, the closing of the sensitive period can result from the down-regulation of mechanisms that support, or the upregulation of mechanisms that restrict, axonal growth. In the visual cortex of cats, for example, levels of the growth-associated protein GAP-43, which is thought to be necessary for axonal growth, decrease precipitously during the critical period. A host of other molecular mechanisms could contribute to a decline in plasticity, including the loss of responsiveness of presynaptic axonal arbors to neurotrophins secreted by postsynaptic neurons, the stabilization of synapses by the extracellular matrix or by proteoglycans, the myelination of axons (preventing them from growing or retracting), the expression of molecules that restrict growth, and the downregulation of molecules that promote growth.

Another major factor that decreases the plasticity of circuits is the experience-driven sharpening of functional tuning. Initially in development, neuronal responses are relatively weak and broadly tuned. Experience causes selective changes in anatomical connections and synaptic efficacy that refine the patterns of both excitatory and inhibitory connections. These changes are self-reinforcing due to the action of self-organizational mechanisms that operate by Hebbian principles (Chapter 46). As a result, once a circuit has been shaped to process information in a certain way, it becomes far more difficult for altered experience to induce new patterns of connectivity.

Deprivation Prolongs Sensitive Periods

Sensitive periods typically close once an animal receives adequate experience. Therefore, when an animal is deprived of appropriate experience, the sensitive period is prolonged. For example, raising songbirds in acoustic isolation prolongs the sensitive period for song memorization, and raising cats or mice in complete darkness prolongs the critical period for binocular processing in the visual cortex. This characteristic of sensitive periods indicates that the event that triggers sensitive period adjustments is the powerful and repeated activation of neurons at the site where changes take place. Without the vigorous activation of these neurons, which requires that both inhibitory as well as excitatory circuitry is effective and balanced, the pathway remains in an uncommitted state and capable of adjusting in response to experience when it becomes available. This characteristic suggests that for young animals (including humans) suffering from a peripheral or central abnormality, a total absence of relevant input is far better than abnormal input. This implies that the optimal therapeutic strategy for such individuals is to deprive them of relevant sensory input until the abnormality is corrected. Otherwise, abnormal sensory experience may close the sensitive period, resulting in a commitment to an abnormal pattern of connectivity that cannot later be reversed.

Reopening Sensitive Periods

The possibility of reopening sensitive periods later in life is of particular importance for therapeutic reasons. For example, by reopening the critical period for binocular

processing in the visual cortex, effective binocular processing and stereoscopic vision might be provided to people who did not develop these capacities due to visual impairment or abnormal eye alignment during the visual critical period. To reinstate sensitive period plasticity completely, the factors (listed in the previous sections) that impose the functional constraints on the relevant circuit must, once again, enable modification by experience. In some circuits, the necessary plasticity may be enabled by the manipulation of only one or a few mechanisms. In the visual cortex, ocular dominance plasticity has been at least partially reinstated in adult animals by, for example, injecting fetal astrocytes or inhibitory interneurons into the cortex, local enzymatic degradation of the extracellular matrix, providing molecules that promote synaptogenesis, blocking the action of molecules that prevent synaptic remodeling, or chronic light deprivation. Interventions like these, when combined with appropriate experience, may have the potential to instate normal function to the affected circuit. With increased understanding of the specific neural circuits and computations that depend on sensitive period learning and the mechanisms that govern their plasticity, acquisition of typical behavior in adult animals that have experienced deprived conditions during sensitive periods may be possible.

Summary

Early experience shapes the functional properties, architecture, and biochemistry of many circuits so that the properties of these circuits are appropriate for the needs and the environment of the individual animal. The changes in circuit properties that occur during a sensitive period differ quantitatively, if not qualitatively, from those that occur in mature circuits. First, they occur readily only during a restricted period in the lifetime of the animal. Second, they involve the selection of particular circuit properties from a wide range of possible properties. Third, the changes in circuit properties that occur during sensitive periods typically do not require the attention of the animal. Fourth, the changes persist throughout life.

Sensitive periods vary in timing and duration across pathways and across species; some last only a few hours, whereas others last until the individual reaches sexual maturity. Sensitive periods open once the information conveyed to a circuit is sufficiently precise and the circuit is competent to process the information and to undergo plastic change. The signal that induces change is probably the repeated, vigorous activation of postsynaptic neurons by presynaptic activity representing the occurrence of an appropriate stimulus or the execution of adaptive behavior. The range of stimuli that are effective in driving plastic change is specified by genetic preprogramming, with most pathways biased heavily to prefer normal patterns of stimulation.

After a sensitive period closes, equivalent conditions have much less effect on the architecture, biochemistry, and functional properties of a circuit. The closure of a sensitive period may be triggered by experience itself or as a consequence of circuit maturation. The mechanisms that close a sensitive period probably vary for different circuits. When experience-induced changes require anatomical remodeling, the end of the sensitive period may be controlled by factors that regulate neurite growth and retraction; when the induced changes require adjustments in synaptic efficacy, the sensitive period will be controlled by factors that influence the capacity of synapses to modify their efficacy; and when the induced changes involve adjustments in the levels of specific receptors, the sensitive period will be controlled by mechanisms that regulate the expression of receptors.

References

Brainard, M. S., & Doupe, A. J. (2000). Auditory feedback in learning and maintenance of vocal behaviour. *Nature Reviews Neuroscience, 1,* 31–40.

DeBello, W. M., Feldman, D. E., & Knudsen, E. I. (2001). Adaptive axonal remodeling in the midbrain auditory space map. *Journal of Neuroscience, 21,* 3161–3174.

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience, 22,* 567–631.

Hensch, T. K. (2005). Critical period plasticity in local cortical circuits. *Nature Reviews Neuroscience, 6,* 877–888.

Hofer, S. B., Mrsic-Flogel, T. D., Bonjoeffer, T., & Hubener, M. (2006). Prior experience enhances plasticity in adult visual cortex. *Nature Neuroscience, 9,* 127–132.

Hubel, D. H., & Wiesel, T. N. (1970). The period of susceptibility to the physiological effects of unilateral eye closure in kittens. *The Journal of Physiology, 206,* 419–436. (London).

Hubel, D., Wiesel, T., & LeVay, S. (1977). Plasticity of ocular dominance columns in the monkey striate cortex. *Philosophical Transactions of the Royal Society of London, Series A, 278,* 377–409.

Immelmann, K. (1972). Sexual imprinting in birds. *Advances in the Study of Behavior, 4,* 147–174.

Knudsen, E. I. (1999). Mechanisms of experience-dependent plasticity in the auditory localization pathway of the barn owl. *Journal of Comparative Physiology A, 185,* 305–321.

Knudsen, E. I. (2002). Instructed learning in the auditory localization pathway of the barn owl. *Nature, 417,* 322–328.

Knudsen, E. I. (2004). Sensitive periods in the development of the brain and behavior. *Journal of Cognitive Neuroscience, 16,* 1412–1425.

Knudsen, E. I., Esterly, S. D., & Olsen, J. F. (1994). Adaptive plasticity of the auditory space map in the optic tectum of adult and baby barn owls in response to external ear modification. *Journal of Neurophysiology, 71,* 79–94.

Konishi, M. (1985). Birdsong: From behavior to neuron. *Annual Review of Neuroscience, 8,* 125–170.

Lein, E. S., & Shatz, C. J. (2001). Neurotrophins and refinement of visual circuitry. In W. M. Cowan, T. C. Sudhog & C. F. Stevens (Eds.), *Synapses* (pp. 613–649). Baltimore, MD: Johns Hopkins Univ. Press.

Miyata, S., Komatsu, Y., Yoshimura, Y., Taya, C., & Kitagawa, H. (2012). Persistent cortical plasticity by upregulation of chondroitin 6-sulfation. *Nature Neuroscience, 15,* 414–422.

Mooney, R. (2009). Neural mechanisms for learned birdsong. *Learn Memory, 16,* 655–669.

Olson, C. R., & Freeman, R. D. (1980). Profile of the sensitive period for monocular deprivation in kittens. *Experimental Brain Research, 39,* 17–21.

Sober, S. J., & Brainard, M. S. (2009). Adult birdsong is actively maintained by error correction. *Nature Neuroscience, 12,* 927–931.

Tropea, D., Van Wart, A., & Sur, M. (2009). Molecular mechanisms of experience-dependent plasticity in visual cortex. *Philosophical Transactions of the Royal Society of London B, 364,* 341–355.

Weaver, I. C. G., Cervoni, N., Champagne, F. A., D'Alessio, A. C., Sharma, S., Seckl, J. R., … Meaney, M. J. (2004). Epigenetic programming by maternal behavior. *Nature Neuroscience, 7,* 847–854.

Suggested Readings

Berardi, N., Pizzorusso, T., & Maffei, L. (2000). Critical periods during sensory development. *Current Opinion in Neurobiology, 10,* 138–145.

Huberman, A. D., & McAllister, A. K. (2002). Neurotrophins and visual cortical plasticity. *Progress in Brain Research, 138,* 39–51.

Jampolsky, A. (1978). Unequal visual inputs and strabismus management: a comparison of human and animal strabismus. In *Symposium on Strabismus, Transactions of the New Orleans Academy of Ophthalmology* (Chapter 26, pp. 358–492). St. Louis, MO: Mosby.

Katz, L. C., & Shatz, C. J. (1996). Synaptic activity and the construction of cortical circuits. *Science, 274,* 1133–1138.

Knudsen, E. I., Heckman, J. J., Cameron, J. L., & Shonkoff, J. K. (2006). Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proceedings of the National Academy of Sciences, 103,* 10155–10162.

Leiderman, P. (1981). Human mother-infant social bonding: Is there a sensitive phase? In K. Immelmann, G. W. Barlow, L. Petrinovich & M. Main (Eds.), *Behavioral development* (pp. 454–468). Cambridge: Cambridge Univ. Press.

Lorenz, K. (1937). Imprinting. *Auk, 54,* 245–273.

Scott, J. P. (1962). Critical periods in behavioral development. *Science, 138,* 949–958.

Smith, G. B., Heynen, A. J., & Bear, M. F. (2009). Bidirectional synaptic mechanisms of ocular dominance plasticity in visual cortex. *Philosophical Transactions of the Royal Society of London B, Biological Sciences, 364,* 357–367.

Eric I. Knudsen

This page intentionally left blank

SECTION IV

SENSORY SYSTEMS

This page intentionally left blank

CHAPTER

22

Fundamentals of Sensory Systems

In bringing information about the world to an individual, sensory systems perform a series of common functions. At its most basic, each system responds with some specificity to a stimulus, and each employs specialized cells—the peripheral receptors—to translate the stimulus into a signal that all neurons can use. Because of their physical or chemical specialization, the many types of receptors transduce the energy in light, heat, mechanical, and chemical stimulation into a change in membrane potential. That initial electrical event begins the process by which the central nervous system (CNS) constructs an orderly representation of the body and of things visible, audible, or chemical. To bridge the distance between peripheral transduction and central representation, messages are carried along lines dedicated to telling the CNS what has taken place in the external world and where it has happened. Such precision requires that labor be divided among neurons so that not only different stimulus energies (light vs. mechanical deformation) but also different stimulus qualities (steady indentation vs. high-frequency vibration of the skin) are analyzed by separate groups of neurons.

In addition to their organization along labeled lines, sensory systems perform common types of operations. Foremost among these is the ability of each system to compare events that occur simultaneously at different receptors, a process that serves to bring out the greatest response where the difference in stimulus strength (contrast) is greatest. At late stages in sensory processing, systems make comparisons with past events and with sensations received by other sensory systems. These comparisons are the fundamental bases of perception, recognition, and comprehension.

This chapter gives an overview of the functional attributes and patterns of organization displayed by the auditory, olfactory, somatosensory, gustatory and visual systems; and it outlines the physiological and anatomical principles common to all sensory systems. When variations on a common theme exist, they are discussed with the goal of bringing the general pattern into sharper focus.

SENSATION AND PERCEPTION

The Function of Each Sensory System Is to Provide the CNS with a Representation of the External World

Because of the changes that occur around an individual, each sensory system has the task of providing a constantly updated representation of the external world. Accomplishing this task is no simple feat because it requires a close interaction between ascending or stimulus-driven mechanisms and descending or goal-directed mechanisms. Together these two mechanisms evoke sensations, give rise to perceptions, and activate stored memories to form the basis of conscious experience. Ascending mechanisms begin with the activity of peripheral receptors, which together form an initial neural representation of the external world. Descending mechanisms work to sort out from the large amount of sensory input those events that require immediate attention. In doing so, the descending mechanisms alter ascending inputs in ways that optimize perception.

Perception of a sensory experience can change even though the input remains the same. A classic example is seen in the image of a vase that can also be perceived as two faces, pointed nose to nose (Fig. 22.1). In this case, the image remains the same—the sensory input remains constant—but the perception of what is being viewed changes as the goal of the viewer changes or as his or her attention wanders. Using this example, it is apparent that detection of a stimulus and recognition that an event

© 2013 Elsevier Inc. All rights reserved.

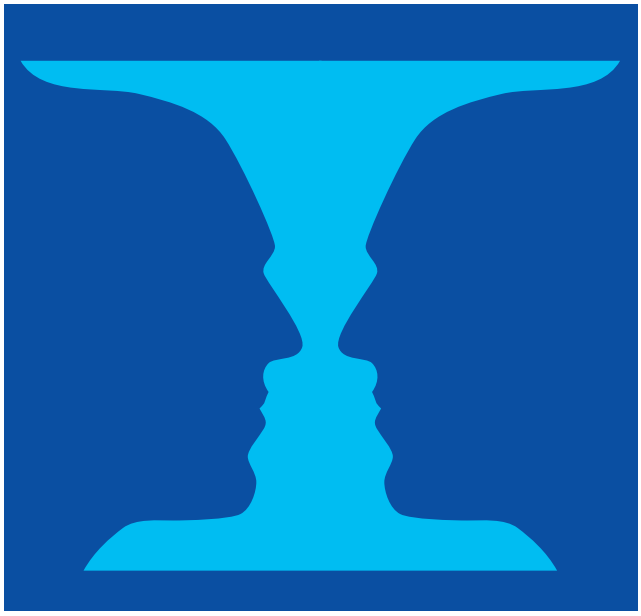


FIGURE 22.1 An example of a figure that can elicit different perceptions (faces or vase) even though stimulus and sensation remain constant. The mind can "see" purple figures against a blue background or a blue figure against a purple background.

has occurred are usually called sensation; interpretation and appreciation of that event constitute perception.

Psychophysics Is the Quantitative Study of Sensory Performance

A psychophysical experiment determines the quantitative relationship between a stimulus and a sensation in order to establish the limits of sensory performance (Stevens, 1957). Such an experiment relies on reports from a subject who is asked to judge quantitatively the presence or magnitude of a stimulus as careful adjustments in the physical attributes are made. One example of threshold detection is the *two-point limen*, in which two blunt probes, separated by a distance that is progressively enlarged or reduced over a series of trials, are applied to the skin surface. The minimum separation distance at which a subject reports two stimuli half the time and one stimulus the other half is taken as the detection threshold. That distance can be measured accurately and is found to vary markedly across the body surface; the two-point limen is smallest for the fingertips and largest for the skin of the back. Other studies, such as those exploring the detection of relative magnitudes of stimuli, can include assessments of object heaviness, loudness of sound, or brightness of light. Studies of this sort have been combined with neurophysiological experiments to compare reports from subjects (sensory behavior) with the responses of single cells (neuronal physiology). Through this procedure the neural mechanisms underlying sensory perception can be examined.

Some general principles hold for all sensation measured in psychophysical experiments. As just pointed out, one principle is that of threshold for detecting a difference between stimuli. Studies look to determine a difference threshold by asking what just noticeable difference (JND) between two stimuli (two lights of different brightness) an observer can detect. E. Weber was the first to formally recognize that small differences between two minimal stimuli are easier to detect than small differences between two robust stimuli. One example is the ability of a subject to detect a difference between two light objects that weigh 0.1 and 0.2 kilograms versus a difference between two heavy objects that weight 10.1 and 10.2 kilograms. The former is much easier than the latter. Weber's law states that the difference threshold for a stimulus is a constant fraction of intensity (the Weber fraction). The formula, $\Delta I/I = k$ describes that law, where I is the intensity of a baseline stimulus, ΔI is the JND between baseline and a second stimulus, and k is the Weber fraction. It is important to recognize that k may be a constant of a particular value for one feature, such as the frequency of sound, but the value changes markedly for another feature of the same sense (sound pressure level). The Weber fraction for sound frequency is exceedingly low (roughly .003), whereas that for sound pressure level is relatively high (0.15). For purposes of comparison, the Weber fraction for luminance is 0.02 and that for concentration of an odorant molecule is 0.10.

G. Fechner proposed that every JND between one stimulus and the next is an equal increment in the magnitude of sensation. That would mean a JND is proportional to a physical variable. His law is formalized in the equation $S = k/\log I$, where S is the sensory experience in terms of magnitude, I is the physically measured intensity of a stimulus, and k is a constant. There is strong intuitive value to this equation because it states the magnitude of a sensory experience is logarithmically related to the physical intensity. Lifting 1 kg and 2 kg produces very different sensory experiences, but lifting 10 kg and 11 kg produces almost the same experience, even though the added weight was equal in the two cases. S. Stevens recognized a century later that rather than a logarithmic relation, perceived sensation and physical intensity were related by a power function, described by the equation $S = kI^P$. Yet the exponent of I could be infinite, depending on the relationship of neural response to stimulus intensity. V. Mountcastle and his colleagues proposed that for mechanosensation the relationship between the physical properties of a stimulus and the response of individual neurons is linear. And most recently, K. Johnson and his colleagues (Johnson et al., 2002) have shown that for the complex percept of roughness perception, a linear relationship exists between subjective experience and neural activity. Thus, a basic law of psychophysics emerges: how an observer perceives a stimulus is a linear function of the intensity of that stimulus.

RECEPTORS

Receptors Are Specific for a Narrow Range of Input

Neurons of the brain and spinal cord do not respond when they are touched or when they are exposed to sound or light or odors. Each form of energy must be transduced by a population of specialized cells, which converts the stimulus into a signal that all neurons understand. In every sensory system, cells that perform this transduction step are called *receptors* (Fig. 22.2). For each of the fundamental types of stimuli (mechanical, chemical, or thermal energy or light), there is a separate population of receptors selective for the particular form of energy. Even within a single sensory system, there are classes of receptors that are particularly sensitive to one stimulus (e.g., heat or cold) and not another (muscle stretch). This specificity in the receptor response is a direct function of differences in receptor structure and chemistry.

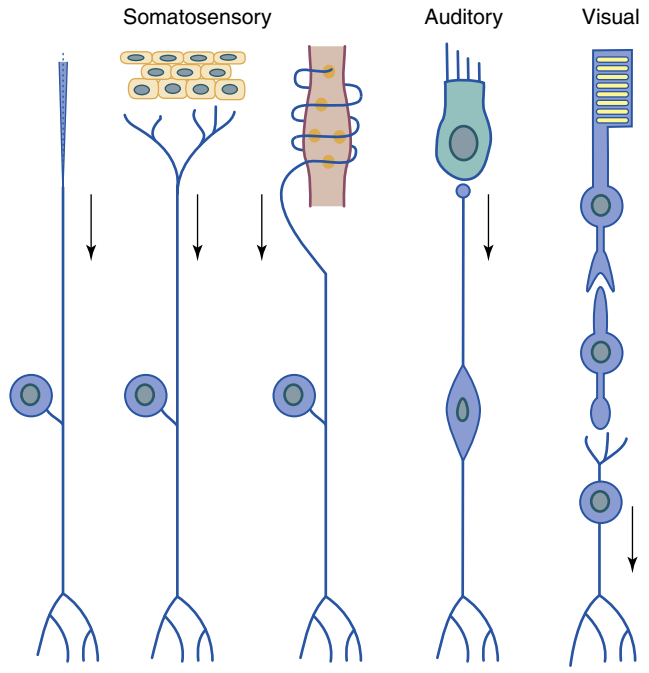


FIGURE 22.2 Receptor morphology and relationship to ganglion cells in the somatosensory, auditory, and visual systems. Receptors are specialized structures that adopt different shapes depending on their function. In the somatosensory system the receptor is a specialized peripheral element that is associated with the peripheral process of a sensory neuron. In the auditory and visual systems, a distinct type of receptor cell is present. In the auditory system, the receptor (hair cell) synapses directly on the ganglion cell, whereas in the visual system, an interneuron receives synapses from the photoreceptor and in turn synapses on the retinal ganglion cell. *Adapted from Bodian (1967).*

Receptor Types Vary across Sensory Systems

Systems differ in the number of distinct receptor types they incorporate, and a correlation exists between the number of receptor types displayed by a system and the types of stimuli that system is able to detect. In the somatosensory system, a large number of receptor types exist to detect many types of stimuli. Separate receptors exist to transduce a variety of mechanical stimuli, including steady indentation of hairless skin, deformation of hair, vibration, increased or decreased skin temperature, tissue destruction, and stretch of muscles or tendons (Fig. 22.2). In the auditory system, two classes of receptor—the inner and outer hair cells of the cochlea—transduce mechanical energy of the basilar membrane, which is set in motion by sound waves (Dallos, 1996). Here, the motility of outer hair cells provides an additional amplification of the basilar membrane motion to increase sensitivity and allow sharp tuning to sound frequency. The inner hair cells respond to the amplified vibrations and excite the large population of neurons upon which they synapse. Thus the two types of receptors act in concert to transduce a single type of stimulus. In the visual system, transduction is performed by two broad classes of receptor in the retina: rods and cones. The number of cone types varies from one in some species to two in many species to three in a few species, but cones in general are tuned to ranges of wavelengths of light. Rods are much more sensitive to light—their intracellular amplification of a single photon's capture generates a much larger current than is achieved in cones—and so they enable vision when light is dim. Olfactory receptors also vary in number from a few hundred types in primates to more than a thousand in rodents. Here the difference between one olfactory sensory neuron (OSN) and the next is a subtle variation on a common theme, since each OSN differs from its neighbor in the primary sequence of a single receptor protein. As that protein varies, so does the OSN's sensitivity to odorant molecules.

Receptors Perform a Common Function in Unique Fashion

All receptors transduce the energy to which they are sensitive into a change in membrane voltage. The task of the receptor is to transmit that voltage change by one route or another to a class of neurons—usually referred to as ganglion cells—that send their axons into the brain or spinal cord (Fig. 22.2). Systems vary in the mechanism whereby receptors and ganglion cells interact. Most receptors in the somatosensory system are part of multicellular organs, the neural components of which are the terminal specializations of dorsal root ganglion cell axons. An appropriate stimulus applied to a somatosensory receptor produces a generator potential—a graded change in membrane voltage—that, when large enough, leads to action potentials

that can be carried over a considerable distance into the central nervous system (CNS). The same approach is used by the olfactory system, since OSNs not only transduce the stimulus of an odorant molecule into a change in membrane potential but also conduct those action potentials into the CNS. We refer to these types of neurons, ones that combine the functions of transduction in the periphery with conduction into the CNS, as long-axon receptors.

Receptor cells of the auditory, vestibular, visual, and gustatory systems are separate, specialized cells that transduce a stimulus and then transmit the resulting signal to the nearby process of a neuron. Because the distances between receptor and target neuron are small, none of these short receptors generate action potentials; they signal their response by a passive flow of current that changes the release of neurotransmitter from either an axon or the base of the cell. These systems differ, however, in the path between receptor and ganglion cell. In the inner ear and in the tongue, sensory receptor cells are modified epithelial cells and not neurons. Nevertheless, they form chemical synapses with the processes of ganglion cells so that these ganglion cells take on the response properties of the receptor cells that innervate them. For these systems, the synapse between a receptor cell and a ganglion cell is little more than the conversion of an analog signal (graded changes in membrane potential) into a digital signal (action potentials). That is not the case in the retina, where photoreceptors relay their response through populations of interneurons interposed between them and retinal ganglion cells (Fig. 22.2). Because of this additional synapse and the opportunity it affords for summation and comparison of receptor signals, the retinal ganglion cell response differs appreciably from that of photoreceptors.

The mechanisms whereby receptor cells transduce and transmit signals are known in greater or lesser detail for each system. Visual transduction is a well-understood, rapid process in which a weak signal (a single photon) can be amplified greatly through a biochemical cascade, leading to the closure of thousands of Na^+ channels and a hyperpolarizing response (Yau & Baylor, 1989). For auditory and vestibular hair cells and somatosensory mechanoreceptors, mechanical deformation of a part of the cell gets transduced into a change in membrane voltage by mechanically gated ion channels, the molecular identity of which is still not known. The response of somatosensory mechanoreceptors is of a single sign because mechanically gated ion channels open and allow depolarizing current to enter (Katz, 1950). This requirement for depolarization may result from the demands placed on the somatosensory ganglion cell to generate action potentials and transmit information over long distances. In contrast, hair cell receptors of the auditory and vestibular systems generate a biphasic response. When protruding villi of a hair cell, called stereocilia, are deflected in one direction, transducer channels open and the cell is depolarized. Yet with deflection of stereocilia in the opposite direction, the same channels close and the cell is hyperpolarized (Hudspeth, 1985). Because sound usually produces a back-and-forth deflection of stereocilia, the result is a back-and-forth movement of the receptor potential—at least for low and moderate frequencies of sound. Thus, the receptor output contains temporal information about the waveform of an acoustic stimulus.

Receptors Have Characteristic Patterns of Position and Density

Receptors are not scattered randomly across the sensory surface. An orderly arrangement of receptors exists along the skin, inner ear, retina, olfactory epithelium, and the lining of the tongue and throat. In the human retina, for example, photoreceptors adopt a hexagonal packing array (Wassle & Boycott, 1991) in the region of highest density, called the fovea. Moreover, only cones are found in the fovea, and for that of humans and other Old World primates, only Red (long wavelength) and Green (middle wavelength) cones are found in the center of the fovea. Hair cells of the vestibular system are even more tightly sequestered, since they occupy very small regions in the semicircular canals and the otolith organs. In the skin, the arrangement of receptors is not nearly so orderly but the density of cutaneous receptors varies markedly across the skin surface. By far the greatest density of receptor terminals is found at the fingertips and the mouth, whereas receptors along the surface of the back are at least an order of magnitude less frequent. In each system, the differences in peripheral innervation density are tightly correlated with spatial acuity. Regions of highest receptor density are also the regions of highest acuity in vision (fovea) and somatic sensation (fingertip).

A perfect test case for innervation density and acuity is seen in the auditory system of microbats. Because these animals use echolocation to navigate and find prey, the auditory system greatly overrepresents the frequencies of sound a bat emits as a probing signal and the surrounding frequencies of Doppler-shifted sounds that echo from objects. Throughout most of the cochlea, the physical properties of the basilar membrane change in a steady fashion so that cochlear hair cells display a progressive shift in the frequency that excites them best. But at the frequencies represented in the Doppler-shifted echo, both the amount of basilar membrane dedicated to those frequencies and the density of inner hair cells along that region increase markedly. The result is a much greater acuity for these information-rich frequencies than for all other frequencies.

Receptors Are the Sites of Convergence and Divergence

The relationship between receptor and ganglion cell is seldom exclusive. Most commonly, a single ganglion

cell receives input from several receptors, and, in many cases, a single receptor sends information to two or more ganglion cells. *Convergence* and *divergence* go hand in hand for the somatosensory system, since a single spot on the skin is often innervated by axons of several ganglion cells, while at the same time the axon of a single ganglion cell can branch to end across a swath of skin. In the somatosensory system, however, the amount of divergence and convergence varies with the class of receptor involved (e.g., thermal receptor vs. mechanoreceptor) and the location of the receptor on the body surface (e.g., shoulder vs. fingertip). Similar features are seen in the mammalian visual system, since divergence and convergence dominate different parts of the retina populated by different receptor types. In the cone-rich central retina, each cone provides as many as five ganglion cells with their main visual drive, whereas in the rod-rich periphery, a few dozen rods supply each ganglion cell with its visual input. In its precision and in its implications for sensory processing, nothing approaches the divergence seen in the cochlea, where a single inner hair cell can be the source of all input received by at least 20 ganglion cells (humans) or as many as 35 ganglion cells (gerbils). Thus, what emerges from a comparison across systems is that convergence and divergence from receptor to ganglion cell vary directly with the demands placed on the system at the specific location. When spatial resolution is a requirement, the convergence of receptor inputs onto individual ganglion cells is low. When detection of weak signals is necessary, convergence is high. When receptor input is used for a complex function or for multiple functions, divergence of input from a single receptor onto many ganglion cells occurs.

Receptors Vary in Their Embryonic Origin

For auditory, vestibular, somatosensory, and olfactory systems, the various classes of receptors and ganglion cells are part of the peripheral nervous system, generated as progeny of neuroblasts located in neural crests and sensory placodes. That is not the case for photoreceptors and retinal ganglion cells. The retina is generated as a protrusion of the embryonic diencephalon and thus all its neurons and supporting cells are CNS derivatives of neural tube origin. As a result of their origin, receptors and ganglion cells of the auditory, vestibular, and somatosensory systems and the OSNs of the nasal epithelium are supported by classes of non-neuronal cells that include modified epithelial supporting cells and Schwann cells. Photoreceptors and retinal ganglion cells, by contrast, are supported by CNS neuroglial cells. Most dramatic of all the consequences resulting from this difference in origin is the ability of axons in somatosensory peripheral nerves to regenerate and reinnervate targets after they are damaged, as opposed to the complete and permanent loss of visual function when optic nerves are cut or crushed.

For all systems except the olfactory, the receptor neurons you were born with are the ones you will live with. Nothing new is added. OSNs, however, have short lives, since they die off and are replaced every 60 days or so. What is seen for all systems, however, is the progressive decline in the number of receptors and in sensory acuity with normal aging. Somatosensory mechanoreceptors in humans are reduced by more than half from the age of 25 years to 65 years. Degeneration of photoreceptors is common as is a progressive reduction in hair cells of both the cochlea and the vestibular organs. Even the OSNs fail to keep up with the ravages of age, since the rate of generation does not match the rate of degeneration. The result in every case is a marked reduction in acuity that can lessen the hedonic value of foods and flowering plants, render a person deaf or blind, and leave him unsteady while standing or walking.

PERIPHERAL ORGANIZATION AND PROCESSING

Sensory Information Is Transmitted along Labeled Lines

A long-appreciated principle that unites structure and function in a sensory system is the doctrine of specific energy, or the *labeled line* principle. This principle states that when a particular population of neurons is active, the conscious perception is of a specific stimulus (Fig. 22.3). For example, activity in one particular population of somatosensory neurons (colored orange in Figure 22.3), leads the CNS to interpret it as a painful stimulus, no matter whether the stimulus is natural (a sharp instrument jabbed into the skin) or artificial (electrical stimulation of the appropriate axons). An entirely separate population of neurons (colored blue in Figure 22.3) would signal light pressure. Why this is so can be seen from the fact that receptors are selective not only in what drives them, but also in the postsynaptic targets with which they communicate. Each ganglion cell transmits its activity into a well-defined region of the CNS, after which a strictly organized series of synaptic connections relays information in a sequence that eventually leads to the thalamus and then to the cerebral cortex (Darian-Smith et al., 1996). It is this orderly relay from receptor to ganglion cell to central neurons at each of several stations that makes up a labeled line. All sensory information arising from a single class of receptors is referred to as a *modality* (e.g., the sensations of pain and light pressure involve distinct modalities). Thus, the existence of labeled lines means that neurons in sensory systems carry specific modalities.

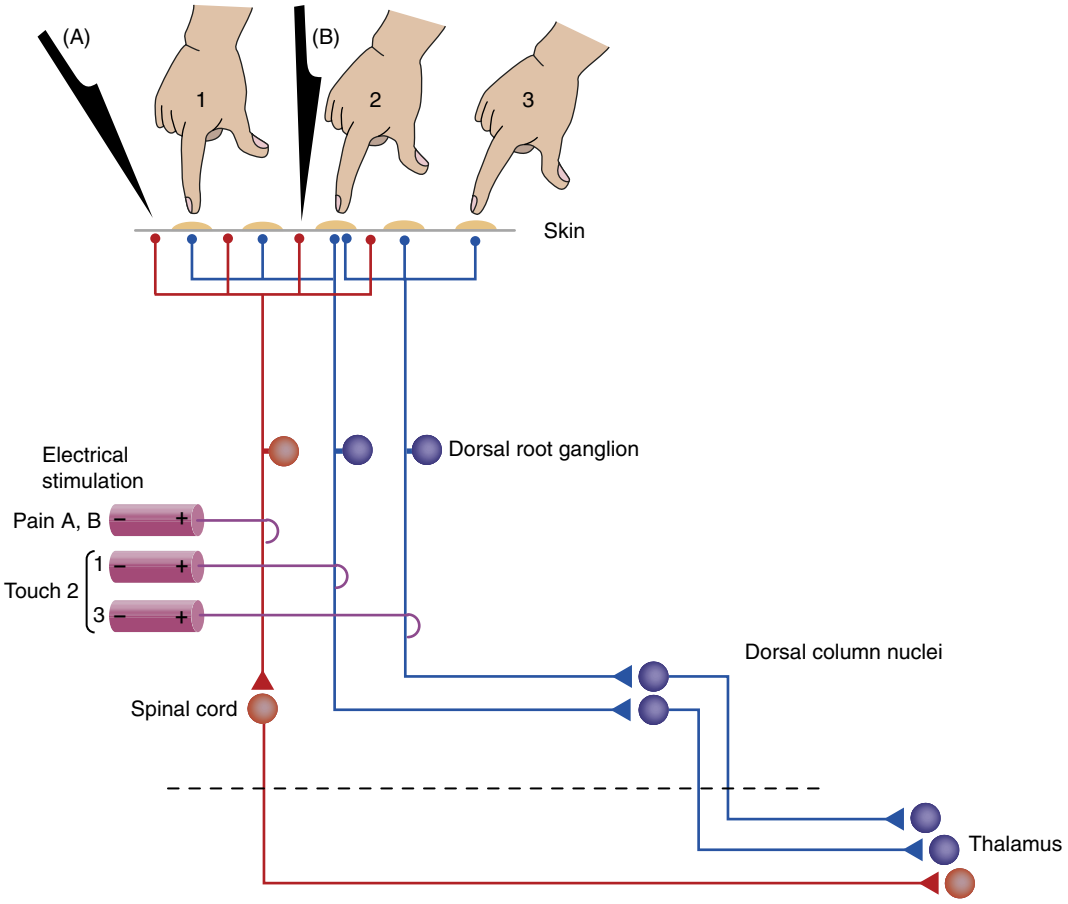


FIGURE 22.3 Example of labeled lines in the somatosensory system. Two dorsal root ganglion (DRG) cells (blue) send peripheral axons to be part of a touch receptor, whereas a third cell (red) is a pain receptor. By activating the neurons of touch receptors, direct touching of the skin or electrical stimulation of an appropriate axon produces the sensation of light touch at a defined location. The small receptive fields of touch receptors in body areas such as the fingertips permit distinguishing the point at which the body is touched (e.g., position 1 vs. position 2). In addition, convergence of two DRG axons onto a single touch receptor on the skin permits touch stimulus 2 to be localized precisely. Electrical stimulation of both axons produces the same sensation, although localized to somewhat different places in the skin. Sharp stimuli (A, B) applied to nearby skin regions selectively activate the third ganglion cell, eliciting the sensation of pain. Electrical stimulation of that ganglion cell or of any cell along that pathway also produces a sensation of pain along that region of skin. Stimuli A and B, however, cannot be localized separately with the pain receptor circuit that is drawn. As the labeled lines project centrally, they cross the midline (decussate) and project to separate centers in the thalamus.

Topographic Projections Dominate the Anatomy and Physiology of Sensory Systems

Receptors in the retina and body surface are organized as two-dimensional sheets, and those of the cochlea form a one-dimensional line along the basilar membrane. Receptors in these organs communicate with ganglion cells and those ganglion cells with central neurons in a strictly ordered fashion, such that relationships with neighbors are maintained throughout. This type of pattern, in which neurons positioned side by side in one region communicate with neurons positioned side-by-side in the next region, is called a *topographic pattern*. As an example, the two touch-sensitive neurons in Figure 22.3 innervate somewhat different positions in the skin. Thus, light touch at position 3 will activate the rightmost ganglion cell in the dorsal root ganglion, whereas touch at position 1 will activate the neighboring ganglion cell. The central projections of these cells are kept separate and activate different targets in the thalamus and above. The end result is a map of the sensory surface of the skin. A different topography is seen in the olfactory system. OSNs divide the nasal epithelium into zones in which sensitivity to a particular range of odorants is maximally represented. A large-scale map has the OSNs in neighboring zones of the epithelium communicate with neighboring regions in the olfactory bulb. Yet the small-scale map from epithelium to bulb displays a fundamental rearrangement in which all OSNs expressing a particular receptor protein send convergent inputs to a single synaptic region (a glomerulus) in the bulb. In general, topographies are a place code for sensory information in which the location of a particular neuron tells you what that neuron responds to, both in place along the sensory surface and in modality.

Neural Signaling Is by a Combination of Rate and Temporal Codes

A great deal of research has been aimed at determining the codes by which neurons signal the presence and the intensity of a stimulus. In addition to place codes, neurons can signal information in the rate at which they respond and in the temporal pattern of their response. For a given receptor, the firing rate or frequency of action potentials signals the strength of the sensory input. The perceived intensity arises from an interaction between this firing rates and the number of neurons activated by a stimulus. Together the number of neurons active with any sensory stimulus and the level of their activity give rise to an intensity code. This is the kind of code used by retinal ganglion cells to signal the intensity (luminance) of light and by spiral ganglion cells to signal the intensity (sound pressure level) of sound. Temporal codes are also used in some systems. For instance, the phase-locking ability of auditory neurons extends to sound frequencies up to a couple thousand cycles per second (kHz), and this code is used for the perception of the pitch of sounds. In addition, all sensory systems must deal with the fact that stimuli can move, as with vibratory stimuli on the skin. This temporal information in a stimulus is carried by the time-varying pattern of activity in small groups of receptors and central neurons.

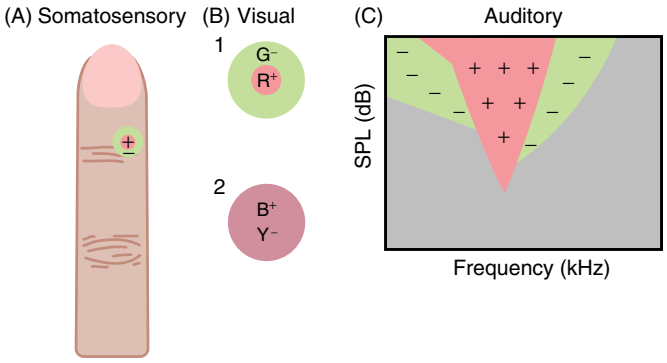


FIGURE 22.4 Center/surround organization of receptive fields is common in sensory systems. In this organization, a stimulus in the center of the receptive field produces one effect, usually excitation, whereas a stimulus in the surround area has the opposite effect, usually inhibition. (A) In the somatosensory system, receptive fields display antagonistic centers and surrounds because of skin mechanics. (B) In the retina and visual thalamus, a common type of receptive field is antagonistic for location and for wavelength. Receptive field 1 is excited by turning on red light (R) at its center and is inhibited by turning on green light (G) in its surround. Receptive field 2 is less common and is antagonistic for wavelength (blue vs. yellow) without being antagonistic for the location of the stimuli. Both are generated by neural processing in the retina. (C) In the auditory system, primary neurons are excited by single tones. The outline of this excitatory area is known as the tuning curve. When the neuron is excited by a tone in this area, the introduction of a second tone in flanking areas usually diminishes the response. This "two-tone suppression" is also generated mechanically, as is seen in motion of the basilar membrane of the cochlea. All of these center/surround organizations serve to sharpen responses over that which would be achieved by excitation alone.

Lateral Mechanisms Enhance Sensitivity to Contrast

A hallmark of all sensory systems is the ability of neurons at even the earliest stages of central processing to integrate the activity of more than one receptor. The most common and easily understood of these mechanisms is *lateral* (or surround) *inhibition* (Fig. 22.4). By this mechanism, a sensory neuron displays a receptive field with an excitatory center and an inhibitory surround (Kuffler, 1953). Such a mechanism serves to enhance contrast: each neuron responds optimally to a stimulus that occupies most of its center but little of its surround. In some cases the comparisons involve receptors of different types so that the center and surround differ not only in sign (excitation vs. inhibition, ON vs. OFF), but also in the stimulus quality to which they respond.

One such example occurs in visual neurons that possess centers and surrounds responsive to stimulation of different types of cones. These cells display a combination of spatial contrast (the difference in the location of cones that produce center and surround) and chromatic contrast (the difference in the visible wavelengths to which these cones respond best). Similar types of responses are evident in the somatosensory system, where the difference between center and surround is the location on the skin from which each is activated. In this case, skin mechanics produce receptive fields with a central hot spot of activity and a surrounding inactive zone. In the auditory system, lateral suppressive areas (Fig. 22.4) are also produced by mechanics, in this case the mechanics of the basilar membrane of the cochlea. These two-tone suppression areas are demonstrated by exciting the auditory nerve fiber with a tone in the central excitatory area and observing the response decrease caused by a second tone in flanking areas (Sachs & Kiang, 1968). In all these sensory systems, center/surround organization serves to sharpen the selectivity of a neuron either for the position of the stimulus or for its exact quality by subtracting responses to stimuli of a general or diffuse nature.

CENTRAL PATHWAYS AND PROCESSING

Axons in Each System Cross the Midline on Their Way to the Thalamus

Axons of ganglion cells entering the CNS form the initial stage in a pathway through the thalamus to the cerebral cortex (Fig. 22.5). Axons in visual, somatosensory, and auditory systems cross the midline—they *decussate*—prior

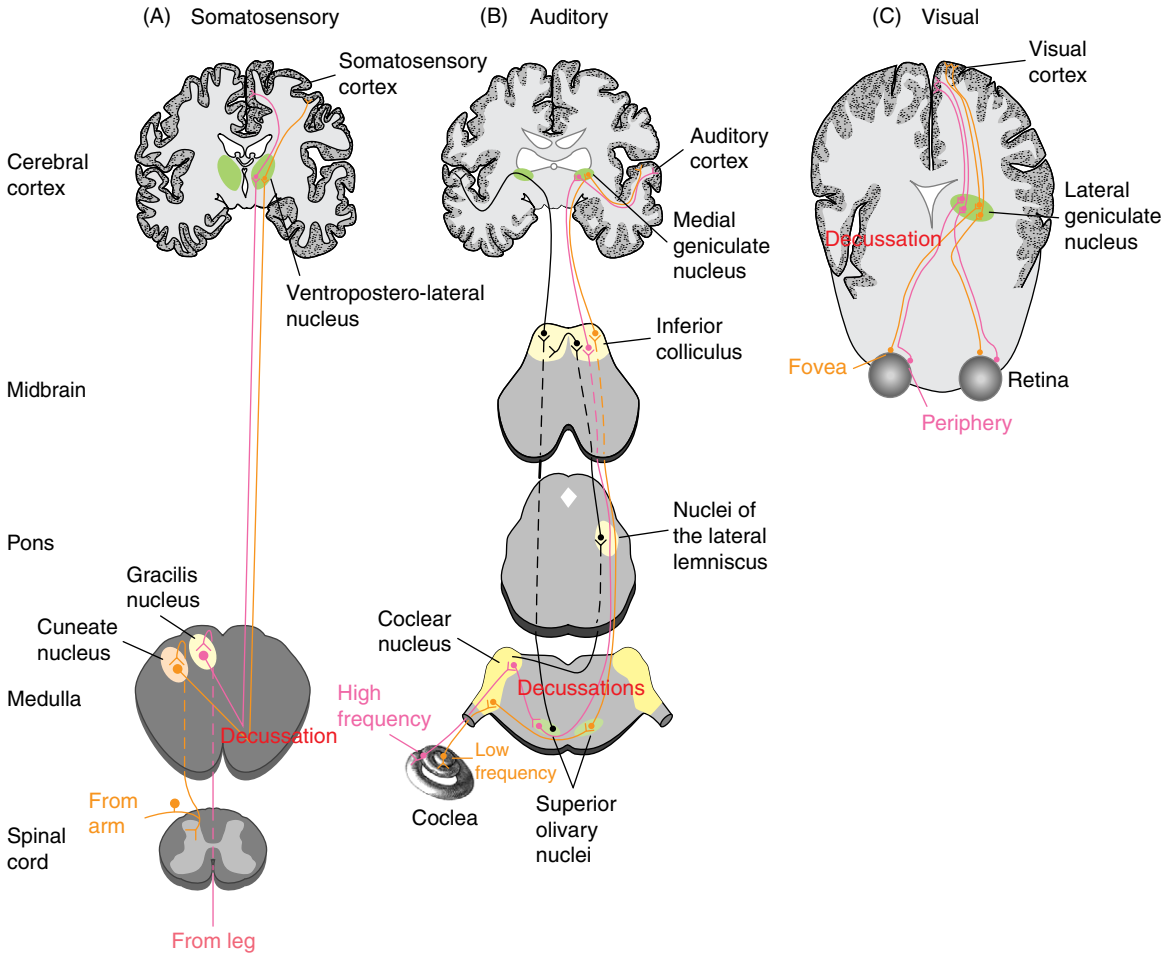


FIGURE 22.5 Comparison of central pathways of sensory systems. In every case, soon after peripheral input arrives in the brain, decussations result in one hemifield being represented primarily by the brain on the opposite side. Each pathway has a unique nucleus in the thalamus and several unique fields in the cerebral cortex. Within each of these areas, the organized mapping that is established by receptors in the periphery is preserved.

to reaching the thalamus, but those of the olfactory and gustatory systems do not. A single, incomplete decussation occurs in animals with frontally placed eyes, where slightly more than half the axons of the optic nerve cross the midline at the optic chiasm. Near complete decussations at the optic chiasm are seen in animals with laterally placed eyes. Decussation in the somatosensory system is nearly total, since all but a small group of axons cross the midline in the spinal cord or brainstem. These decussations serve two broad functions: They bring together into one hemisphere all axons carrying information from half the visual world or they bring somatosensory information into alignment with visual input and motor output. In contrast, multiple decussations occur in the auditory system prior to the thalamus, since comparison of input from the two ears is the dominant requirement of sound source localization. Nevertheless, at high levels in the auditory system, one side of the brain is concerned mainly with processing information about sound sources located toward the opposite side of the body, as demonstrated by lesion/behavioral studies of sound localization.

Specific Thalamic Nuclei Exist for Each Sensory System

Information from all sensory systems except the olfactory are relayed through the thalamus on its way to the cerebral cortex (Fig. 22.5). Olfactory information reaches primary olfactory cortex without a relay in thalamus. Yet even in this sense, perception of odorants and discrimination of one odorant from another occurs only after a thalamic relay. This relay of sensory input through the thalamus involves either a single large nucleus or, in the case of the somatosensory systems, two nuclei: one for the body and one for the face. In each thalamic nucleus, synaptic circuits are said to be secure because activity in presynaptic axons usually leads to a postsynaptic response. Within the

thalamic nuclei, neurons performing one function (e.g., relay of discriminative touch) are segregated from those performing another (e.g., relay of pain and temperature). Even within one function, mappings of neurons are preserved so that there is separation of neurons providing touch information from the arm versus from the leg and of neurons responding to low versus high sound frequencies (Fig. 22.5). Usually, for each thalamic nucleus, there is a population of large neurons and one or two populations of small neurons (Jones, 1981). In each case the larger neurons carry the most rapidly transmitted signals from the periphery to the cortex.

Multiple Maps and Parallel Pathways

Nuclei in the central pathways often contain multiple maps. For instance, in the auditory system, axons of spiral ganglion cells divide into branches as they enter the CNS and terminate in three subdivisions of the cochlear nucleus. Each division contains its own map of sound frequency that was originally established by the cochlea. The greatest number of maps generated from ganglion cell input is found in the visual system of primates, in which as many as 12 separate retinotopic maps are stacked on top of one another in the lateral geniculate nucleus (LGN, in the thalamus), which receives direct input from the retina. Such a large number of distinct maps in the LGN is indicative of inputs from ganglion cells that vary in location, structure, and function. For vision, one idea is that surface features such as color and form are carried along a path separate from the one that handles three-dimensional features of motion and stereopsis. The functional significance of multiple maps in general, however, remains to be clarified. Perhaps the need for multiple parallel paths exists because of the relatively slow speed and the limited capacity of single neurons. So rather than have the same group of neurons perform different functions in serial order, each of several parallel groups performs a separate function. This leads eventually to the problem of binding together all features of a stimulus into a coherent percept, the neural basis for which may be the synchronized activity of neurons across several areas of the cerebral cortex (Singer, 1995).

SENSORY CORTEX

Sensory Cortex Includes Primary and Association Areas

Axons of sensory relay nuclei of the thalamus project to a single area or a collection of neighboring areas of the cerebral cortex, thereby providing them with a precise topographic map of the sensory periphery. These parts of the cortex are frequently referred to as *primary sensory areas* (Fig. 22.6). Neighboring areas with which the primary areas communicate directly or by a single intervening relay area are sensory *association areas.*

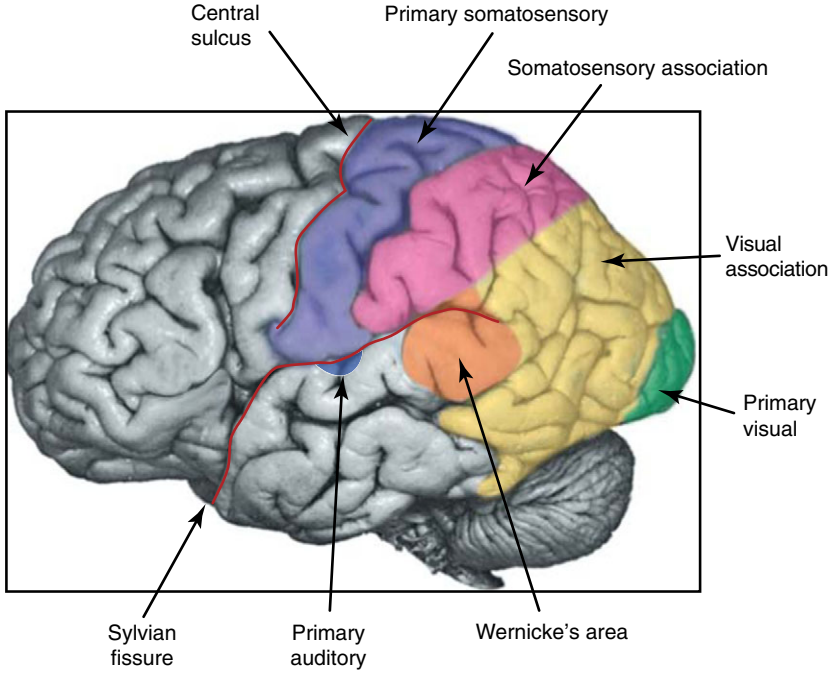


FIGURE 22.6 The location of primary sensory and association areas of the human cerebral cortex. The primary auditory cortex is mostly hidden from view within the Sylvian fissure. *From Guyton and Hall (2000).*

Response Mappings and Plasticity

Each area of sensory cortex shares with its subcortical components a map of at least part of the sensory periphery. Thus, retinotopic, somatotopic, and tonotopic maps are evident in the relevant areas of cortex; no such map exists for the chemical senses, however. The retina and skin are two-dimensional sheets so the map of the sensory periphery on the surface of the cortex is a simple transformation of the peripheral representation onto the cortex. In the auditory periphery, there is a one-dimensional mapping of frequency. This tonotopic mapping is faithfully represented along one dimension of cortical distance, and the orthogonal direction may map a second, as yet undiscovered, property.

As indicated above, the distribution of receptors is uneven for most systems. Thus, the fovea of the primate retina and the fingertips of the primate hand are regions that possess a high density of receptors with small receptive fields. Such an uneven distribution of neurons devoted to a structure is further amplified in the CNS. A much greater percentage of neural machinery subserves the representation of the retinal fovea or the fingertips than deals with the representation of other regions of the retina or body surface (Fig. 22.7). This expansion of a representation in the CNS, referred to as a *magnification factor*, appears particularly impressive in humans, in whom a very large part of primary visual cortex is devoted to the couple of millimeters of retina in and around the fovea. In the auditory cortex, such magnification of frequency representation is again seen in echolating bats, since the region devoted to the Doppler-shifted signaling frequency is much enlarged over that for any other set of frequencies.

It is now clear that maps of sensory cortex are not fixed and immutable but rather plastic. In the somatosensory system, if input from a restricted area of the body surface is removed by severing a nerve or by amputation of a digit, the portion of the cortex that was previously responsive to that region of the body surface becomes responsive to neighboring regions (Merzenich et al., 1984). In the auditory system, following high-frequency hearing loss, the portion of cortex previously responsive to high frequencies becomes responsive to middle frequencies. Frequency tuning of neurons or auditory cortex can also be shifted with classical conditioning methods (Weinberger, 2007). Such plasticity of cortical maps requires some time to be established and could result from strengthening of already established lateral connections or from growth of new connections. It is likely but not firmly established that the same types of mechanisms cause cortical changes during the processes of learning and memory.

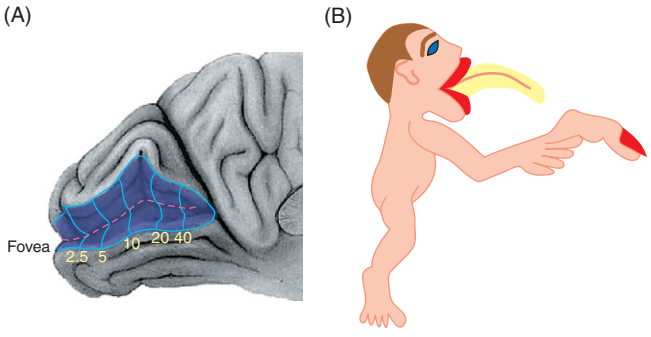


FIGURE 22.7 Examples of sensory magnification in the visual and somatosensory systems. (A) Determination of a visual field map in the human primary visual cortex shows that more than half this area is devoted to the central 10° of the visual field. Very little is devoted to the visual periphery beyond 40°. *From Horton and Hoyt (1991).* (B) Figure of how the human body would appear if the body surface were a perfect reflection of the map in the first somatosensory cortex. The mouth and tongue and the tip of the index finger enjoy a greatly enlarged representation in the thalamus and cortex.

A Common Structure Exists for Sensory Cortex

Neurons in areas of the sensory cortex (and most other areas of the cerebral cortex) are organized into six layers. The middle layers (III and IV) are the main site of termination of axons from the thalamus (Fig. 22.8). In the primary sensory cortices, these middle layers are enlarged and contain many small neurons. Because the small cells resemble grains of sand in standard histological preparations, the sensory areas are themselves referred to as granular areas of cortex.

Columnar Organization

Properties other than place in the periphery are mapped in primary sensory areas of the cortex. The third spatial dimension of the cortex, that of depth, arranges neurons in adjacent 0.5- to 1-mm-wide regions, referred to as *columns* (Mountcastle, 1997). In these columns, neurons stacked above and below one another are fundamentally similar but differ significantly from neurons on either side of them. One example of columns with a clear anatomical correlate is the division of the primary visual cortex of most primates and some carnivores into a series of alternating regions dominated by the right and left retinas. Each ocular dominance column contains cells driven exclusively or predominantly by one eye; adjacent columns are dominated by the other eye. Other properties, such as selectivity for the orientation of a visual stimulus and the contrast between it and the surround, are also arranged in columns of primary visual cortex (Hubel, 1988). Similar types of columns are evident in the somatosensory system as regions of modality and place specificity. They are most clearly seen in the representation of mystacial vibrissae in rodent somatosensory cortex, where a one-to-one matching of vibrissa with cortical territory gives rise to structures called barrels. In the auditory

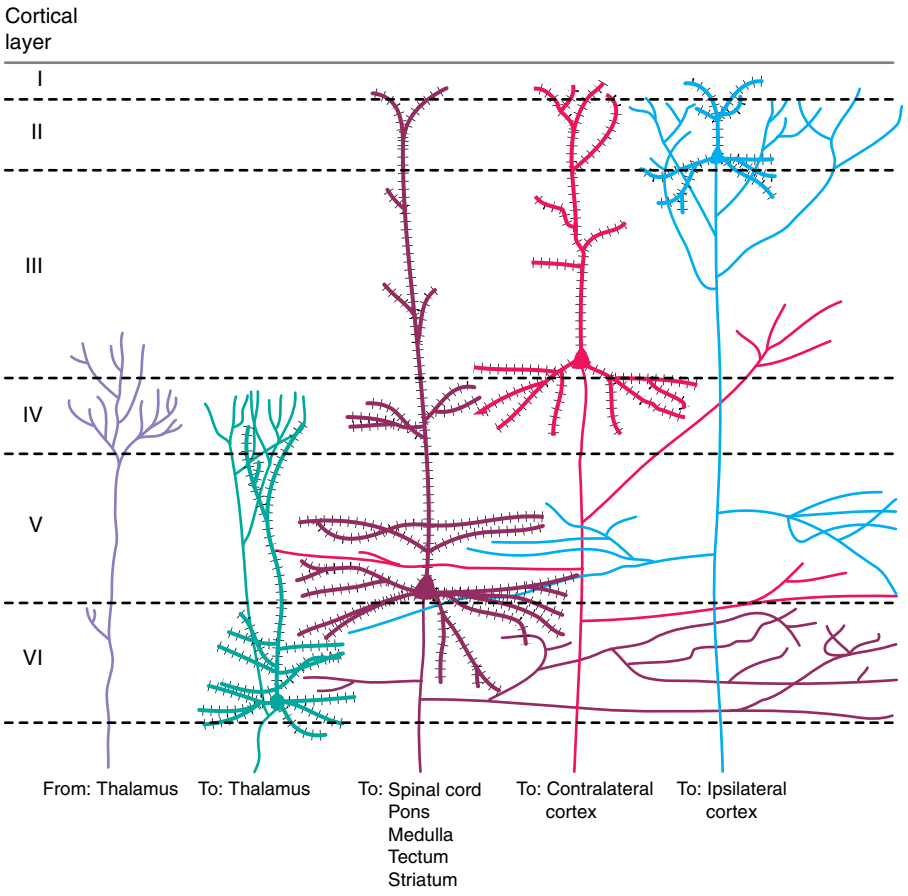


FIGURE 22.8 Cellular organization of the sensory cortex into six layers (I–VI). Inputs from the thalamus terminate mainly in layers III and IV. The main output neurons of the cortex are pyramidal cells, which are distributed in different layers according to their projections. Descending projections are to the thalamus (neurons in layer VI) or to the spinal cord, pons and medulla, tectum, or striatum (neurons at various levels in layer V). Ascending projections to other "higher" cortical centers are often from neurons located above layer IV; some of these projections are to the same hemisphere (layer II) or to the opposite hemisphere (layer III). *Adapted from Jones (1984).*

system, neurons within a column generally share the same best frequency and the same type of binaural interaction characteristic: either one ear excites the neurons and the other inhibits or suppresses the response to the first ear (suppression column) or one ear excites and the other ear also excites or facilitates the response to the first ear (summation column). Moreover, in areas of nonprimary cortex, the feature displayed most commonly by neurons of a particular area is one that often comes to occupy columns. A good example is found in the middle temporal area (MT or V5) of visual association cortex, where neurons are tuned for the direction of a moving visual stimulus. Neurons selective for one particular direction of visual stimulus movement are organized into columns through the depth of MT; these are flanked by columns of neurons tuned for other directions of movement (Albright, Desimone, & Gross, 1984). So consistent are these findings among sensory, motor, and association areas that columnar organization is viewed as a principal organizing feature for all of the cerebral cortex (Mountcastle, 1997).

Stereotyped Connections Exist for Areas of Sensory Cortex

Neurons of the cerebral cortex send axons to subcortical regions throughout the neuraxis and to other areas of the cortex (Fig. 22.8). Subcortical projections are to those nuclei in the thalamus and brainstem that provide ascending sensory information. By far the most prominent of these is to the thalamus: the neurons of a primary sensory cortex project back to the same thalamic nucleus that provides input to the cortex. This system of descending connections is truly impressive because the number of descending corticothalamic axons greatly exceeds the number of ascending thalamocortical axons. These connections permit a particular sensory cortex to control the activity of the very neurons that relay information to it. One role for descending control of thalamic and brainstem centers is likely to be the focusing of activity so that relay neurons most activated by a sensory stimulus are more strongly driven and those in surrounding less well activated regions are further suppressed.

A large number of cortical neurons project to other areas of cortex (Fig. 22.8). Cortico-cortical projections link primary and association areas of the sensory cortex and establish parallel paths so that different aspects of vision, audition, and somatic sensation come to be handled by different areas of cortex. These connections establish a hierarchy within a system, such that "ascending" or "forward" connections begin with neurons from superficial cortical layers (I–III) and end with axonal terminations mainly in layers III and IV of higher cortical regions. Similarly, the ascending projection from the thalamus terminates mainly in these layers in primary sensory cortices. Corresponding descending projections from higher to lower cortical regions begin in the deep or superficial cortical layers and project to layers outside of III and IV. In addition to projections to the ipsilateral hemisphere of cortex, there are also projections to the contralateral hemisphere via the corpus callosum and other commissures. In visual and somatosensory systems, these commissural connections are restricted in origin and termination; they exist to unite the representation of midline structures into a coherent percept (Hubel, 1988).

Response Complexity of Cortical Neurons

Responses of cortical neurons in primary sensory cortices are more complex than those seen for neurons in the periphery. One example is seen in the primary visual area of the cerebral cortex, where neurons are responsive to stimuli that are not concentric circles (center and surround) but elongated lines possessing a specific orientation. Comparable synthesis of simpler inputs to reconstruct more complex features of stimulus is apparent at higher levels in the visual system (Logothetis & Sheinberg, 1996) and in the somatosensory and auditory areas of the cerebral cortex.

Physiologically, processing and selectivity for stimulus features become progressively more complex within the hierarchically organized pathways that connect primary with association areas of the cortex (Gallant & Van Essen, 1994). In the visual system, separate "streams" involved in visuosensory and eventually visuomotor functions have been described; one is responsible for using visual cues to drive appropriate eye movements and the other for dealing with the tasks of visual perception (Gallant & Van Essen, 1994). In the somatosensory system, separate motor and limbic paths exist to perform much the same functions for the entire body, supplying sensory input to coordinate and adjust motor output and using complex input from many receptor types to match the shape of a tactual stimulus with one already stored in memory (Johnson & Hsiao, 1992). In the association pathways of the human auditory system, a specialized area of cortex, Wernicke's area, plays a fundamental role in processing speech and language information and in communicating with Broca's area to form a speech motor response. These streams are not separate because traditionally viewed "motor" areas such as Broca's are now known to become activated in comprehension tasks. Apparent from this pattern in association areas of the cortex is the continued pressure for a division of labor within each sensory system, not one that produces separate paths for analyzing elemental features of a stimulus but one that combines those features either to elicit appropriate movements or to match a stimulus with an internal representation of the world.

SUMMARY

The functional organization of sensory systems shares common themes of transduction, relay, organized mappings, parallel processing, and central modification. It is no surprise that a case has been made for a common phylogenetic origin of sensory systems. Differences among the systems, however, demonstrate that each has existed and operated independently for as long as there have been vertebrates. What remains in overview is a well-ordered basic plan from periphery to perception that has been modified in its details as variations in niche have led to specializations in function.

References

Albright, T. D., Desimone, R., & Gross, C. G. (1984). Columnar organization of directionally selective cells in visual area MT of the macaque. *Journal of Neurophysiology, 51,* 16–31.

Bodian, D. (1967). Neurons, circuits and neuroglia. In G. C. Quarton, T. Melnechuk, & F. O. Schmitt (Eds.), *The neurosciences: A study program* (pp. 6–24). New York, NY: The Rockefeller Press.

Dallos, P. (1996). Overview: Cochlear neurobiology. In P. Dallos, A. N. Popper & R. R. Fay (Eds.), *The cochlea* (pp. 1–43). New York, NY: Springer-Verlag.

Darian-Smith, I., Galea, M. P., Darian-Smith, C., Sugitani, M., Tan, A., & Burman, K. (1996). The anatomy of manual dexterity: The new connectivity of the primate sensorimotor thalamus and cerebral cortex. *Advances in Anatomy, Embryology, and Cell Biology, 133,* 1–142.

Gallant, J. L., & Van Essen, D. C. (1994). Neural mechanisms of form and motion processing in the primate visual system. *Neuron, 13,* 1–10.

Guyton, J. E., & Hall, A. C. (2000). *Textbook of medical physiology,* (10th ed.). New York, NY: W. B. Saunders and Company.

Horton, J. C., & Hoyt, W. F. (1991). The representation of the visual field in human striate cortex: A revision of the classic Holmes map. *Archives of Ophthalmology, 109,* 816–824.

Hubel, D. H. (1988). *Eye, Brain and Vision*. New York, NY: Freeman.

Hudspeth, A. J. (1985). The cellular basis of hearing: The biophysics of hair cells. *Science, 230,* 745–752.

Johnson, K. O., & Hsiao, S. S. (1992). Tactual form and texture perception. *Annual Review of Neuroscience, 15,* 227–250.

Johnson, K. O., Hsiao, S. S., & Yoshioka, T. (2002). Neural coding and the basic law of psychophysics. *The Neuroscientist, 6,* 111–121.

Jones, E. G. (1981). Functional subdivision and synaptic organization of the mammalian thalamus. *International Review of Physiology, 25,* 173–245.

Jones, E. G. (1984). Laminar distribution of cortical efferent cells. In A. Peters & E. G. Jones (Eds.), *Cerebral Cortex* (Volume 1, pp. 521–553). New York, NY: Plenum.

Katz, B. (1950). Depolarization of sensory terminals and the initiation of impulses in the muscle spindle. *Journal of Physiology* (London), *111,* 261–282.

Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology, 16,* 37–68.

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience, 19,* 577–621.

Merzenich, M. M., Nelson, R. J., Stryker, M. P., Cynader, M. S., Schoppmann, A., & Zook, M. (1984). Somatosensory cortical map changes following digit amputation in adult monkeys. *Journal of Comparative Neurology, 224,* 591–605.

Mountcastle, V. B. (1997). The columnar organization of the neocortex. *Brain, 120,* 701–722.

Sachs, M. B., & Kiang, N. Y. S. (1968). Two-tone inhibition in auditory-nerve fibers. *Journal of the Acoustical Society of America, 43,* 1120–1128.

Singer, W. (1995). Time as coding space in neocortical processing: A hypothesis. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 91–104). Cambridge, MA: MIT Press.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64,* 153–181.

Wassle, H., & Boycott, B. B. (1991). Functional architecture of the mammalian retina. *Physiological Review, 71,* 447–480.

Weinberger, N. M. (2007). Auditory associative memory and representational plasticity in primary auditory cortex. *Hearing Research, 70,* 226–251.

Yau, K. -W., & Baylor, D. A. (1989). Cyclic GMP-activated conductance of retinal photoreceptor cells. *Annual Review of Neuroscience, 12,* 289–328.

Suggested Readings

Beurg, M., Fettiplace, R., Nam, J. -H., & Ricci, A. (2009). Localization of inner hair cell mechanotransducer channel using high-speed calcium imaging. *Nature Neuroscience, 12,* 553–558.

Dowling, J. E. (1987). *The retina: An approachable part of the brain*. Cambridge, MA: Belknap Press.

Stewart H. Hendry and Steven S. Hsiao

This page intentionally left blank

CHAPTER

23

Chemical Senses: Taste and Olfaction

Some of the most remarkable feats in the animal kingdom are accomplished using chemical detection. For example, the male silk moth follows the scent of a female for several miles to reach its mate. A great white shark is able to detect one drop of blood in an Olympic-size swimming pool to identify prey. More than 30 million chemical compounds have been catalogued. Different chemical combinations emitted by foods, predators, and mates allow animals to distinguish among them. The gustatory and olfactory senses allow detection of this enormous number of chemicals and translate this information into meaningful behaviors. The identification of large chemoreceptor families, the ability to monitor cell activity in the periphery and central nervous system, and the ability to genetically manipulate chemosensory systems in model organisms are revolutionizing our understanding of chemosensory systems and explaining how the brain encodes flavors and fragrances.

TASTE

The sense of taste is primarily involved in feeding, allowing animals to identify food that is nutrient-rich and to avoid toxic substances. The sensory cues detected by the gustatory system are soluble chemicals, limiting detection to a short range by direct contact with a chemical source. The concentration range for taste detection is broad and depends on the nature of the chemical stimulus. At one extreme, taste cells detect sugars and amino acids at very high concentrations (100 millimolar), allowing animals to detect food with high nutritional value. At the other extreme, taste cells can also detect minute amounts of noxious substances or toxins, compounds that are harmful at very low concentrations. Mammals are thought to perceive only five taste modalities: sweet, bitter, sour, salty, and umami (the taste of the amino acid glutamate), with several chemicals in most categories.

How the brain translates chemical detection into the perception of different taste modalities and taste behaviors is a basic problem in neural coding. Of course, the gustatory system does not work in isolation, and the smell, texture, and sight of food also contribute to flavor perception. For example, without our olfactory neurons, we are unable to recognize coffee, chocolate, or wine! However, it is the gustatory system that acts as the final checkpoint controlling food acceptance or rejection and is essential for the basic recognition of taste modalities. Understanding the neural coding of taste information begins with knowledge about how taste receptor cells detect chemical cues, an area of rapid advances. How taste quality is encoded in higher brain areas is an area of active investigation that will ultimately require unraveling neural circuits and neural activity underlying taste behavior.

Taste Cells Are Situated within Taste Buds Located in Several Distinct Subpopulations

In mammals, taste recognition begins on the tongue, where specialized cells detect chemical cues. Mammalian taste cells are derived from epithelial tissue rather than neural tissue, although they display neural properties such as depolarization and release of neurotransmitters. The apical tip of each receptor cell contains microvilli that project into the mucus of the oral environment. The microvilli are the site of taste detection, where taste ligands interact with membrane-bound receptors or ion channels. The basolateral membrane of taste cells forms chemical

© 2013 Elsevier Inc. All rights reserved.

synapses with primary gustatory nerve fibers. The nerve fibers enter the base of the taste bud and transmit taste detection to the brain.

Groups of 50 to 100 taste cells are clustered together into taste buds (Fig. 23.1). The apical center of the taste bud is the taste pore, where the microvilli of taste cells interact with the environment. Groups of taste buds are located primarily within papillae on the tongue, although they are distributed on other parts of the oral cavity as well (Fig. 23.2). Three classes of papillae are defined based on their location on the lingual epithelium: fungiform papillae on the anterior tongue, foliate papillae on the posterior sides, and circumvallate papillae on the posterior midline. In addition, taste cells are found on the palate and the epiglottis. Taste receptor cells undergo constant renewal, arising from an underlying population of basal epithelial cells.

Taste buds are innervated by dendrites of neurons that travel in the facial (VIIth), glossopharyngeal (IXth), or vagus (X) nerves to gustatory nuclei in the brainstem. Different regions of the tongue are innervated by different nerves. Fungiform papillae and the foliate papillae on the anterior tongue are innervated by the chorda tympani branch of the facial nerve. Taste buds on the soft palate receive fibers from the greater superficial petrosal (GSP) branch of the facial nerve. Circumvallate and posterior foliate papillae are innervated by the lingual-tonsillar branch of the glossopharyngeal nerve. Taste buds of the epiglottis and the esophagus are innervated by the internal branch of the superior laryngeal nerve (SLN), which is a branch of the vagus (Xth) nerve. A single nerve fiber may innervate cells in more than one taste bud, and each taste bud is innervated by several different afferent fibers.

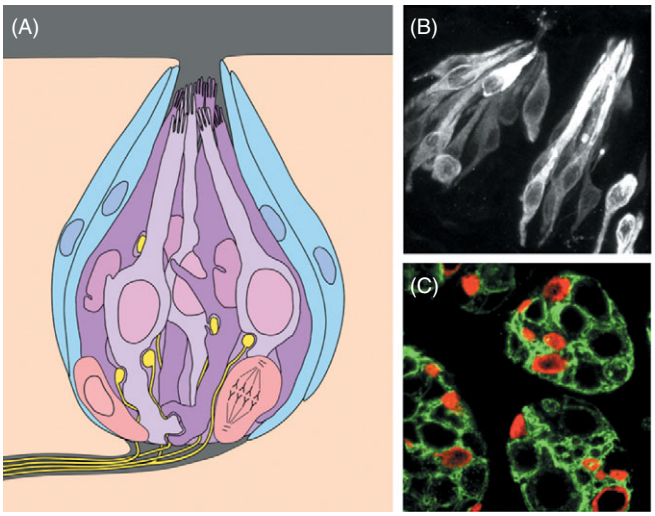


FIGURE 23.1 Cell types in mammalian taste buds. (A) The taste bud contains approximately 50–100 taste cells. These epithelial receptor cells make synaptic contact with distal processes of cranial nerves VII, IX, or X, whose cell bodies lie within the cranial nerve ganglia. Microvilli of the taste receptor cells project into an opening in the epithelium, the taste pore, where they make contact with gustatory stimuli. (B) The characteristic spindle shape of taste receptor cells is revealed when a subset of taste cells is labeled with an antibody against α-gustducin, a gustatory G protein. (C) Gustducin-positive cells appear round in cross section, as shown by α-gustducin immunoreactivity (red), whereas other cells are revealed with an antibody against the H blood group antigen (green).

Mechanisms of Taste Detection by Receptor Cells

Recent fundamental discoveries have begun to identify the receptor molecules that detect taste ligands. There are two different families of G-protein-coupled receptors that mediate the detection of sugars: amino acids and bitter compounds. In addition, candidate ion channels for the recognition of salts and sour compounds have been identified.

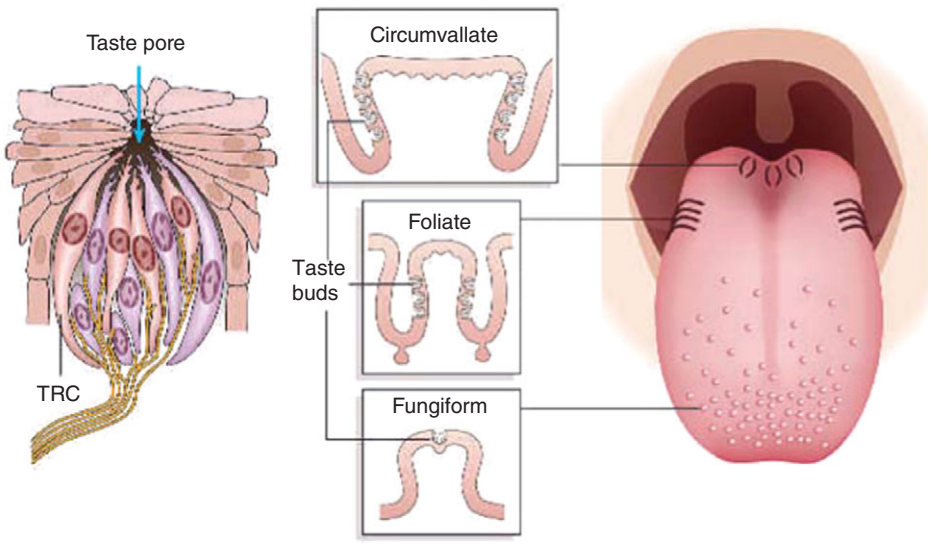


FIGURE 23.2 Diagram of a tongue showing the distribution of various taste bud populations, which are found in the fungiform papillae on the anterior tongue, the vallate and foliate papillae on the posterior tongue. These taste buds are innervated by branches of the VIIth, IXth, and Xth cranial nerves (see text). *Reprinted by permission from Macmillan Publishers LTD:* Nature 444*(7117): 288–94, copyright 2006.*

The T1R Receptor Family Mediates the Taste of Sugars and Amino Acids

In mammals, a small family containing three genes mediates the detection of sugars and amino acids. The Taste Receptor 1 family is comprised of the T1R1, T1R2, and T1R3 receptors (Hoon et al., 1999). These receptors are G-protein-coupled receptors with large extracellular domains similar to the metabotropic glutamate receptors. The receptors function as heterodimers. T1R2 plus T1R3 recognizes sugars and T1R1 plus T1R3 detects amino acids (Nelson et al., 2001, 2002).

The following evidence demonstrates that the combination of T1R2 plus T1R3 recognizes sugars. First, as expected for taste receptors, the T1Rs are expressed in taste cells on the tongue. Second, mice engineered to lack either T1R2 or T1R3 do not detect sugars, as determined by electrophysiological recordings of taste nerves (Fig. 23.3) and by behavioral taste tests. Whereas normal mice prefer sugar water to water, mice lacking these receptors do not. Third, heterologous cells that coexpress T1R2 and T1R3 respond to sugars, but those that express T1R2 or T1R3 alone do not. This argues that T1R2 and T1R3 act together to detect sugars. Equivalent experiments demonstrate that the combination of T1R1 and T1R3 recognizes amino acids. Thus, T1R1 is selective for amino acids, T1R2 sugars, and T1R3 is an obligate partner for both amino acid and sugar detection.

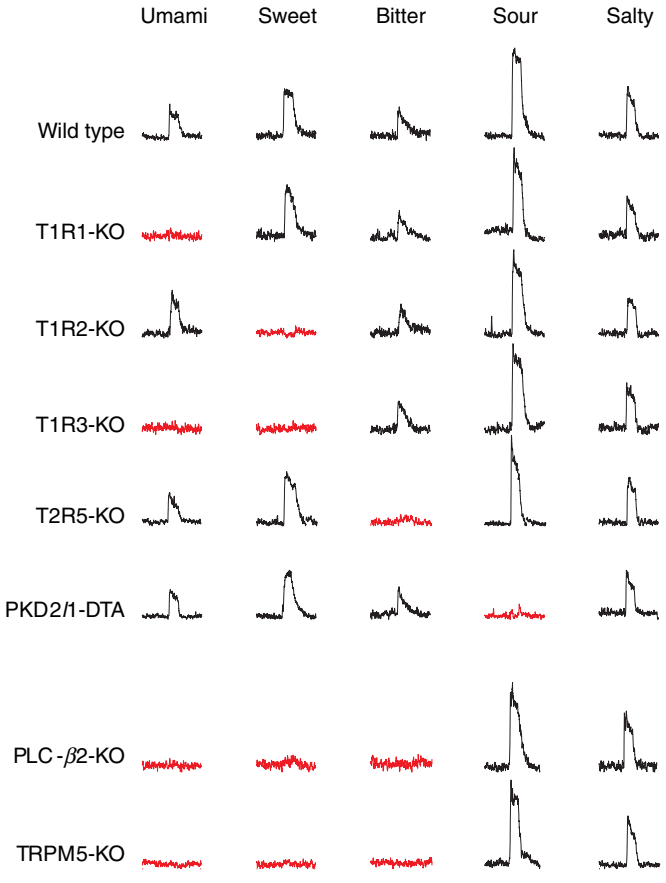


FIGURE 23.3 Mice lacking specific taste receptors have specific taste defects. Shown are nerve recordings showing taste-induced activity to amino acids, sugars, bitter compounds, sour compounds, and salts. Wild-type mice respond to all compounds. Mice lacking T1R1 do not detect amino acids, mice lacking T1R2 do not detect sugars, and mice lacking T2R5 do not detect specific bitter compounds. *Reprinted by permission from Macmillan Publishers LTD:* Nature 444*(7117): 288–94, copyright 2006.*

Mammals detect a wide array of sugars and amino acids only at high concentrations (10–100 mM), allowing them to recognize compounds that provide nutritional value. The T1R receptors have apparently solved the problem of detection by having low affinity ligand-binding sites. Differences in ligand recognition are found in different species. For example, the taste of umami in humans is the taste of the amino acid glutamate. The human T1R1 plus T1R3 amino acid receptor binds to glutamate with a tenfold higher affinity than other amino acids. In contrast, the mouse T1R1 plus T1R3 receptor recognizes all amino acids with similar affinity. Thus, mice have an amino acid taste, whereas humans have an umami taste. Although this may sound like humans are short-changed, it may not be the case. Because foods that have one amino acid generally contain many, the enhanced ability to detect glutamate in humans may actually enhance detection of all amino acids. The observation that T1R sequence variation underlies differences in taste detection also occurs for sugars. For example, we find aspartame (Nutrasweet) sweet, but mice and rats do not detect it due to amino acid differences in T1Rs. Interestingly, cats lack a functional T1R2 receptor, providing a plausible explanation for why they do not taste sugars. These studies highlight the role of T1Rs in sugar and amino acid taste detection in mammals and demonstrate that sequence differences in T1Rs can account for species-specific taste preferences.

The T2R Receptor Family Mediates Bitter Taste Detection

A common variation in human behavior is the ability to detect the bitter compound phenylthiocarbamide (PTC). In 1931, a scientist at the Dupont chemical company synthesized this compound and quickly came to the realization that although it was flavorless to him, his colleague found it unbearably bitter. Since then, numerous human taste tests (often done in high school classes) have revealed that the ability to detect PTC and a related compound, n-propyl uracil (PROP), behaves as a classic inherited recessive trait with taster and nontaster alleles (see Box 23.1). The identification of the genomic location linked to PTC detection enabled the identification of genes associated with this chromosomal interval and the seminal discovery of a large family of related mammalian bitter taste receptors.

The Taste Receptor 2 family (T2R) of bitter receptors is a family of G-protein-coupled receptors whose members, unlike T1Rs, contain a small extracellular domain, and most

BOX 23.1

IDENTIFYING A GENETIC LOCUS LINKED TO BITTER TASTE PERCEPTION

The ability to taste the compound phenylthiocarbamide (PTC) varies dramatically among humans. Approximately 75% of people find PTC intensely bitter, whereas 25% do not detect it. The ability to detect PTC was linked to a small region of chromosome 7 in human genetic studies. The taste receptor gene T2R38 is located within this interval, and variations in its sequence perfectly correlate with the ability to taste PTC. In the human population, there are two major sequence variations in T2R38, which differ by three amino acids. One variant (containing the amino acid substitutions AVI) accounts for the nontasters and the other variant (containing the amino acid substitutions PAV) is the major taster allele. Thus, people with two AVI alleles are nontasters, those with one PAV and one AVI are tasters, and those with two PAV alleles are super-tasters (Kim et al., 2003). These studies demonstrate that humans vary in their ability to detect taste compounds because of variations in the sequence of taste receptors.

Kristin Scott

resemble the opsin family of GPCRs (Adler et al., 2000). There are 25 functional T2R receptors in humans, 35 in mouse and rat, and 3 in chicken. Most T2R genes are linked in large chromosomal arrays, suggesting rapid expansion of this gene family.

A number of lines of evidence conclusively demonstrate the T2Rs are bitter receptors. First, T2Rs are specifically expressed in taste cells on the tongue. Second, the receptors are localized to taste cell dendrites decorating the taste pore, the site of taste detection. Third, heterologous cells expressing T2Rs respond to bitter compounds; for example, cells expressing mT2R5 respond specifically to cyclohexamide, whereas cells containing mT2R8 respond strongly to denatonium and weakly to PROP. Fourth, mice lacking specific T2Rs show specific bitter taste defects (Mueller et al., 2005) (Fig. 23.3).

Studies of the T2R receptor family provide important insights into how animals recognize a large number of bitter compounds and are able to detect them at very low concentrations. Each T2R receptor generally recognizes only a few bitter compounds with very high affinity, allowing for very specific detection. In addition, the large number of T2R receptors ensures that animals can recognize several bitter compounds.

Mechanisms of Taste Transduction for T1Rs and T2Rs

Activation of the T1R and T2R receptors leads to cell depolarization and neurotransmitter release. In general, signal transduction mediated by G-protein-coupled receptors involves a conformational change in the receptor leading to activation of a heterotrimeric G protein, in turn activating enzymes whose byproducts regulate the activity of ion channels. For T1Rs and T2Rs, the signaling cascade proposed is that activation of T1Rs or T2Rs leads to activation of gustducin, a heterotrimeric G protein, leading to activation of the enzyme phospholipase C-β2 (PLC-β2) and depolarization by the ion channel TRPM5 (Fig. 23.4). These signaling molecules are all specific to taste cells, arguing that they are specialized for taste transduction. Importantly, mice lacking gustducin, PLC-β2 or TRPM5 do not taste sugars or bitter compounds.

Ion Channels Sense Sour and Acid Taste

Unlike sweet and bitter taste detection that rely on activation of G-protein-coupled receptors and downstream signaling molecules to regulate channel activity, the tastes of sour and salty are likely to be directly mediated by ion channels.

Sour taste is produced by acids, and the degree of sourness depends primarily on proton concentration. Patch-clamp studies suggest that several different ion channels may participate in sour transduction, which is not surprising because protons are capable of modulating most ion channels. A recent study has identified the cells that detect sour taste and express cation channels named PKD1L3 and PKD2L1 (Huang et al., 2006). PKD1L3 and PKD2L1 are cationic ion channels of the TRP family (transient receptor potential) of ion channels. Transgenic mice have been produced such that the PKD2L1 promoter drives expression of a toxin gene to kill all cells that express the channel. Mice lacking these cells do not show electrophysiological responses to sour compounds, but still respond to sugars, salts, and bitter compounds. These experiments argue that cells containing PKD2L1 are necessary for sour detection.

The ion channel that mediates sodium salt taste is the epithelial sodium channel ENaC. This channel, which is responsible for sodium transport in a variety of epithelial tissues, is expressed on the apical membrane of salt-sensitive taste cells, where it mediates the passive influx of sodium. Sodium simply diffuses through the open channels to depolarize taste cells; and is presumably pumped out by a NaK-ATPase on the basolateral membrane. The first

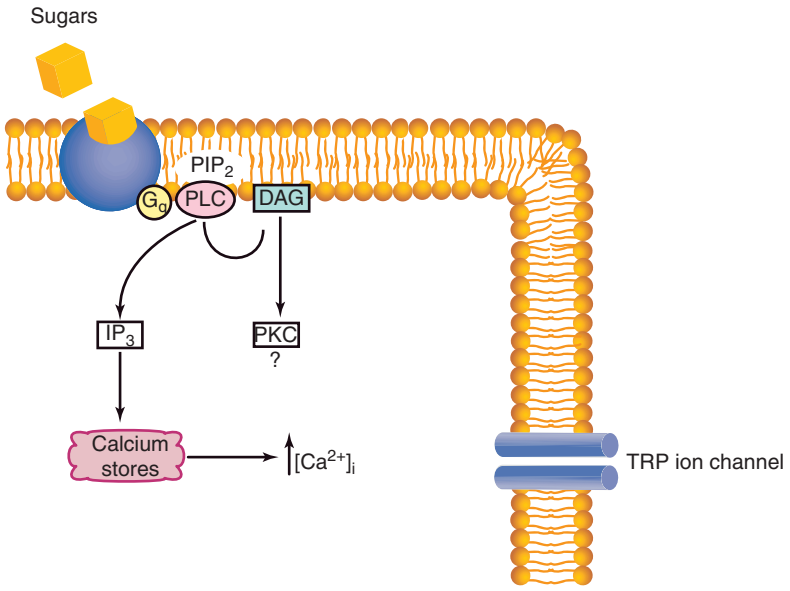


FIGURE 23.4 Taste transduction for sugars, amino acids, and bitter compounds uses a common signaling pathway. Activation of the receptor activates a heterotrimeric G protein. This activates phospholipase C-β2 (PLC-β2), producing the second messengers IP_3 and diacylglycerol (DAG). Activation of PLC-β2 leads to opening of the TRPM5 ion channel and cell depolarization.

evidence for a role of ENaC in taste came from experiments showing that the gustatory nerve response to NaCl was inhibited by amiloride, a diuretic drug known to block these channels in other transporting epithelial tissues. More recently, animals lacking ENaC only in taste cells have been shown to have specific defects in salt taste detection (Chandrashekar et al., 2010).

The Organization of Taste Receptor Genes on the Tongue

Taste receptors are segregated into different taste cells on the tongue. *In situ* hybridization experiments as well as immunohistochemistry have been used to determine the expression patterns of receptor mRNAs and proteins. These experiments demonstrated that T1Rs and T2Rs are not found in the same taste cells (Fig. 23.5). Among the T1Rs, T1R1 plus T1R3 and T1R2 plus T1R3 are segregated into different cells. In addition, the PKD1L3 and PKD2L1 candidate sour ion channel subunits are found in different cells than the T1Rs or T2Rs. This demonstrates that there are modality-specific taste cells on the tongue, with one taste cell population recognizing sugars, another amino acids (or glutamate for humans), a third bitter compounds, and a fourth sour cues. ENaC is expressed in a fifth cell population. An important conclusion is that different taste modalities are encoded by the activation of different taste cells. Our perception of only a few different tastes may stem from the fact that there are only a few different types of taste cells on the tongue. Sugar-sensing and bitter-sensing cells are also found in *Drosophila melanogaster*, suggesting that modality-specific cells are a common principle of taste systems (see Box 23.2).

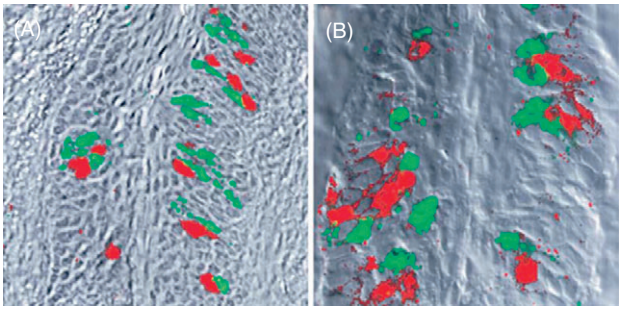


FIGURE 23.5 Different taste cells express different taste receptor genes. *In situ* hybridization experiments showing expression of T1Rs (green) and T2Rs (red) in mammalian tongue sections. *From Nelson et al. (2001).*

Activation of Taste Neurons Is Sufficient to Generate Taste Behavior

The finding that different taste compounds activate different populations of taste cells suggests that activation of different cells in the periphery leads to different taste percepts and different taste behaviors. To test this, transgenic mice were engineered to express novel receptors in taste cells. In an elegant experiment, transgenic mice were engineered to contain the human bitter receptor (hT2R16) for phenyl-β-D-glucopyranoside (a compound that humans perceive as bitter and mice do not detect) in either T2R (bitter) or T1R2 (sugar) cells (Fig. 23.6). If hT2R16 is expressed in T2R cells, mice avoid phenyl-β-D-glucopyranoside. If this human bitter receptor is expressed in T1R2 cells, however, mice prefer phenyl-β-D-glucopyranoside. Thus, expression of a bitter receptor

BOX 23.2

TASTE CELLS IN FRUIT FLIES

Drosophila melanogaster, like mammals, recognize sugars, salts, and bitter compounds, making them a model system for comparative studies of taste recognition. *Drosophila* taste with sensory neurons on their proboscis (mouth), legs, and wings. A family of 68 candidate Gustatory Receptor genes, unrelated to mammalian T1R or T2R receptors, may mediate taste detection because some members are expressed in taste tissue and recognize taste ligands in heterologous systems. Similar to mammalian taste, multiple receptors are expressed in each *Drosophila* taste cell. Two different classes of taste cells have been identified based on the receptor subsets they contain. One population detects bitter compounds and a second population detects sugars. Additional subpopulations of taste neurons likely recognize other taste categories. Thus, although flies and mammals are separated by more than 500 million years of evolution, they both have taste cells that selectively recognize bitter compounds or sugars. This suggests that modality-specific taste cells may be a general strategy that organisms use to distinguish nutrients from toxins.

Kristin Scott

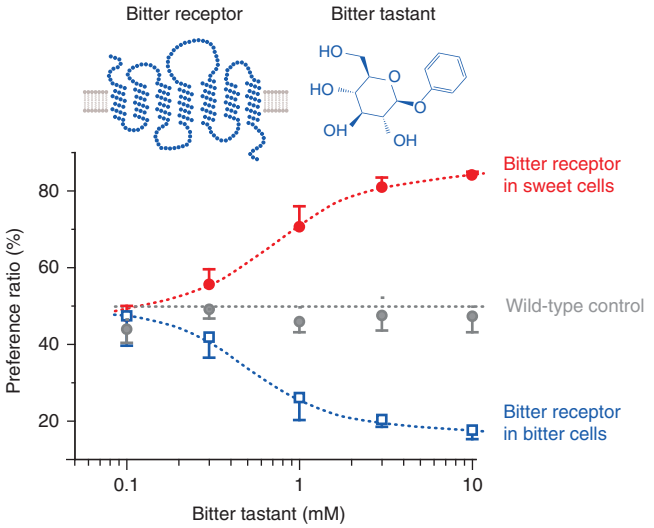


FIGURE 23.6 Activation of different taste cells is tethered to specific taste behaviors. Mice were engineered to contain a human bitter receptor in either T1R2 sugar sensing or T2R bitter sensing cells. Mice avoid the specific bitter compound when the receptor is expressed in T2R cells and prefer the same compound when the receptor is expressed in T1R2 cells. This demonstrates that the same compound can elicit different behaviors depending on which cells detect it. *Reprinted by permission from Macmillan Publishers LTD:* Nature 444*(7117): 288–94, copyright 2006.*

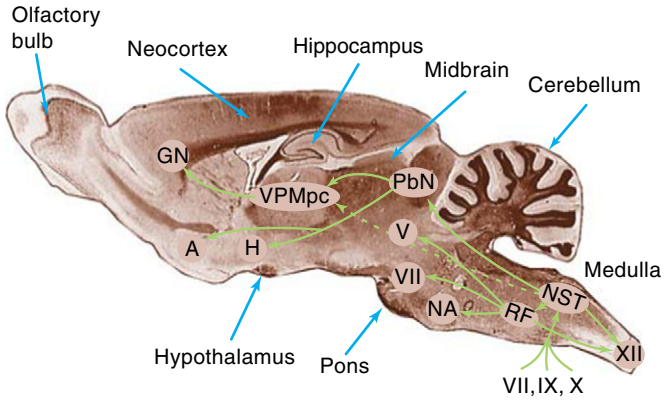


FIGURE 23.7 Schematic diagram of the ascending gustatory pathway; descending projections are not shown. Connections of the rodent gustatory system within the CNS are shown by solid lines; the projection from NST to VPMpc in primates is indicated by a dashed line. NST, nucleus of the solitary tract; PbN, parabrachial nuclei; VPMpc, ventero-posteromedial nucleus (parvi cellularis) of the thalamus; GN, gustatory neocortex; A, amygdala; H, hypothalamus; NA, nucleus ambiguus; RF, reticular formation; V, VII, and XII, trigeminal, facial, and hypoglossal motor nuclei; VII, IX, and X, axons of peripheral gustatory fibers in the facial, glossopharyngeal, and vagal cranial nerves.

in sugar cells triggers attraction to the bitter compound. These experiments demonstrate that activation of sugar cells leads to taste acceptance behavior and activation of bitter cells leads to avoidance. This argues that the activity of different taste cells is hardwired to different behaviors.

Transmission of Taste Information from the Tongue to the Brain

The three nerves that contact taste cells have cell bodies in ganglia and send axons to the solitary tract nucleus (NST) of the medulla in a topographic order (Fig 23.7). Dendrites from the chorda tympani (a branch of the facial nerve VII) contact the anterior tongue, cell bodies reside in the geniculate ganglion, and axons terminate in the rostral pole of the NST. Dendrites from the lingual-tonsillar branch (glossopharyngeal nerve IX) contact the posterior tongue, have cell bodies in the petrosal ganglion, and have axons that invade the NST at intermediate regions. Dendrites from the superior laryngeal branch (vagus nerve X) innervate the epiglottis and esophagus, have cell bodies in the nodose ganglion, and send axons to the caudal regions of the NST. Thus, information from the anterior-posterior tongue is represented from the rostral to caudal NST.

This segregation continues throughout the gustatory pathway to the cortex, where there are separate terminal fields for VIIth and IXth nerve inputs. From the NST, ascending fibers project in most species to third-order cells within the parabrachial nuclei (PbN) of the pons. A thalamocortical projection arises from the PbN to carry taste information to the parvicellular portion of the ventroposteromedial nucleus of the thalamus (VPMpc) and on to the gustatory neocortex (GN), located in rodents within the agranular insular cortex. In primates, taste fibers bypass the pontine relay and project directly to the VPMpc.

Arising in parallel with the thalamocortical projection is a second projection that carries gustatory afferent information into limbic forebrain areas involved in feeding and autonomic regulation, including the lateral hypothalamus, the central nucleus of the amygdala, and the bed nucleus of the stria terminalis. Descending axons within the gustatory system arise from the insular cortex and several ventral forebrain areas and project to the PbN and NST. There are also numerous local connections among neurons within the NST and with cells of the oral, facial, and pharyngeal motor nuclei (V, VII, ambiguous, and XII), either directly (as with XII) or via interneurons in the reticular formation. These hindbrain systems form the substrate for many taste-mediated somatic and visceral responses related to ingestion and rejection of tastants.

Models of Encoding Taste Quality

The nature of the neural coding of taste quality has been debated vigorously for many years, with considerable disagreement about whether taste quality is represented by activity in specific neural channels (a labeled line code) or by the relative activity across the responsive neurons (a population code) (Fig. 23.8). The labeled line model of taste coding proposes that different taste qualities are encoded by the activation of different cells. In this model, cells respond to selective cues in the periphery, and this information remains segregated in the brain. This hypothesis suggests that "sweetness" is coded by activity of neurons that respond selectively to sugars, "saltiness" by activity of neurons that respond selectively to salt, and so forth. Activity in a given cell type provides complete information about the quality of the stimulus.

In the population coding model of taste quality (across-fiber patterning), each cell responds to multiple taste modalities with different activity levels for different taste ligands, such that taste quality is coded by the pattern of activity across taste fibers. In this coding hypothesis, the pattern of activity generated across the entire array of taste neurons encodes taste quality, whereas activity in any one cell cannot unambiguously represent both stimulus quality and intensity. This population approach to quality coding makes the multiple sensitivity of gustatory neurons an essential part of the neural code for taste quality; it stresses that the code for quality is given in the response of the entire population of cells, placing little or no emphasis on the role of an individual neuron. In this scenario, the animal then discerns different tastes by differences in the ensemble activity of taste neurons.

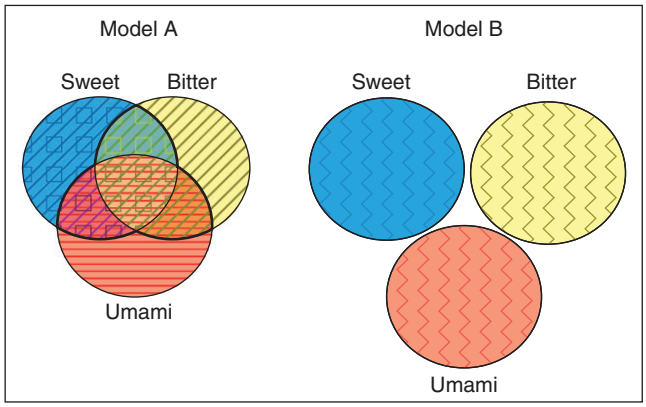


FIGURE 23.8 Two models of taste coding have been proposed. The left panel depicts the labeled line model in which different taste cells recognize different taste qualities, such that one cell population is activated by sugars and different cells are activated by bitter compounds. The right panel depicts the population coding model in which cells respond to multiple taste modalities. The pattern of activity across the ensemble of taste neurons encodes for specific tastes. *From Amrein and Bray (2003).*

Taste Coding in the Periphery

The studies of taste receptor genes strongly argue for the labeled line encoding of taste information in the periphery. Different taste cells contain different taste receptor genes and respond to different taste ligands. T1R2 plus T1R3 cells detect sugars, T1R1 plus T1R3 cells detect amino acids, T2R cells detect bitter compounds, and PKD2L1 cells detect acids. There is no overlap in the expression of different receptor types and thus no cell that detects both sugar and bitter compounds. Thus, "sweetness" is encoded by the activation of sugar cells, and "bitter" by the activation of bitter-sensing cells.

Not only do different cells identify different tastes, but these different cell types are necessary and sufficient for taste detection and behavior. Mice lacking T1R2 do not detect sugars, but still detect bitter, salt, sour, and amino acids. Similarly, mice lacking cells containing PKD2L1 do not detect sour, but do detect sugar, bitter, salt, and amino acids. Moreover, simply activating the sugar cells artificially generates taste acceptance behavior and activating the bitter cells generates avoidance behavior. These experiments are difficult to reconcile with population coding models. In population coding, the relative activity of nonselective neurons dictates taste quality, and loss of any subpopulation of taste cells would affect detection of all

compounds. In addition, in population coding, activation of a subset of taste cells would only be meaningful in the context of the ensemble activity. Instead, different taste cells recognize different taste modalities and mediate specific taste behaviors, arguing that there are labeled lines of taste information from peripheral activation to behavior.

Taste Coding in the Central Nervous System

Although the selectivity of taste cells in the periphery strongly argues for labeled lines of taste information, neurophysiological recordings of both primary gustatory nerves and central gustatory neurons typically show that individual fibers or cells respond to more than one of the stimuli representing the salty, sweet, sour, or bitter taste qualities, often to as many as three or four. Accumulating evidence suggests that there are functional classes of neurons that correspond in some way to primary taste qualities—that is, "sucrose-best" cells and "sodium-best" cells. However, analyses show that no single class of neurons in isolation can discriminate well between different taste qualities. As a result, an "across neuron pattern" theory suggesting that taste quality is coded by the relative activity across a population of neurons was proposed.

The current debate in the mammalian taste field is how to reconcile the evidence of labeled lines for different tastes in the periphery with the evidence for mixed lines in the central nervous system. One possibility is that a single nerve fiber synapses onto taste cells of many different taste qualities on the tongue, such that the labeled lines in the periphery become mixed at the first synapse. Alternatively, electrophysiological studies of single neurons with defined connections to taste cells, rather than random recordings of neurons in the solitary tract nucleus or gustatory cortex, may reveal populations of higher-order taste neurons that show selective responses to single taste qualities, supporting labeled line encoding.

Summary

The sense of taste allows animals to recognize different chemical compounds and gives rise to a limited number of taste sensations (saltiness, sweetness, sourness, bitterness, and umami/amino acids). Two families of taste receptors, the T1R and T2R families, mediate the detections of sugars, amino acids, and bitter compounds. These receptors are G-protein-coupled receptors that activate a heterotrimeric G protein, phospholipase C, and TRPM5, leading to cell depolarization and neurotransmitter release. Ion channels have been implicated in the detection of salts and sour compounds.

The transduction machinery for taste detection is found in receptor cells within taste buds distributed in several subpopulations, innervated by different peripheral nerves. These nerves project into the nucleus of the solitary tract in the medulla. From there, projections arise to the parabrachial nuclei in the pons and then to the thalamus and gustatory neocortex. A parallel pathway carries taste information into the ventral forebrain to areas involved in autonomic regulation. How gustatory information is encoded in the brain is an area of active investigation. In the periphery, an individual taste cell is tuned to a single taste quality, arguing that there are labeled lines of taste information. More centrally, gustatory neurons appear broadly tuned to stimuli of different taste qualities, leading to a debate whether the labeled lines in the periphery become mixed centrally. The ability to determine the connectivity of higher order taste neurons and their taste response profiles will enhance our understanding of taste neural circuits and the coding mechanisms underlying taste perception.

OLFACTION

The number of volatile cues in the environment that animals can detect is colossal. Humans, although not known to have particularly keen noses, can sense over 10,000 different odors, with detection thresholds in the parts per million and even parts per billion. Scent-tracking dogs, especially bloodhounds, have even more sensitive noses, approximately 10 million times more sensitive than humans, allowing them to follow scent trails a few days old. The ability to detect chemicals a distance from the source allows animals to direct their movement toward food and potential mates and away from harmful or dangerous environments. Although many animals rely on their olfactory system for survival, the sense of smell in humans is primarily an aesthetic sense, enhancing our enjoyment of foods and evoking memories of past experiences.

How does the olfactory system allow for the recognition and discrimination of thousands of different odors? A large family of 1,000 olfactory receptors provides the molecular diversity necessary for recognizing numerous ligands. The discovery of the olfactory receptor genes has provided molecular tools to probe the molecular underpinnings of olfactory detection. The logic of olfactory detection in the periphery is discussed next.

Odor Molecules Are Transduced by Olfactory Receptor Neurons

The nasal epithelium is the site for odor detection in mammals (Fig. 23.9). Olfactory sensory neurons have ciliated endings that terminate in the mucosal layer exposed to the environment. Axons of olfactory neurons cross the cribiform plate and synapse in the olfactory bulb, the most rostral region of the brain. Olfactory receptor molecules on the ciliated endings detect odors, causing a change in neural activity that is transmitted to the brain. The detection of odors by sensory cells and the organization of synaptic connections provide insight into the

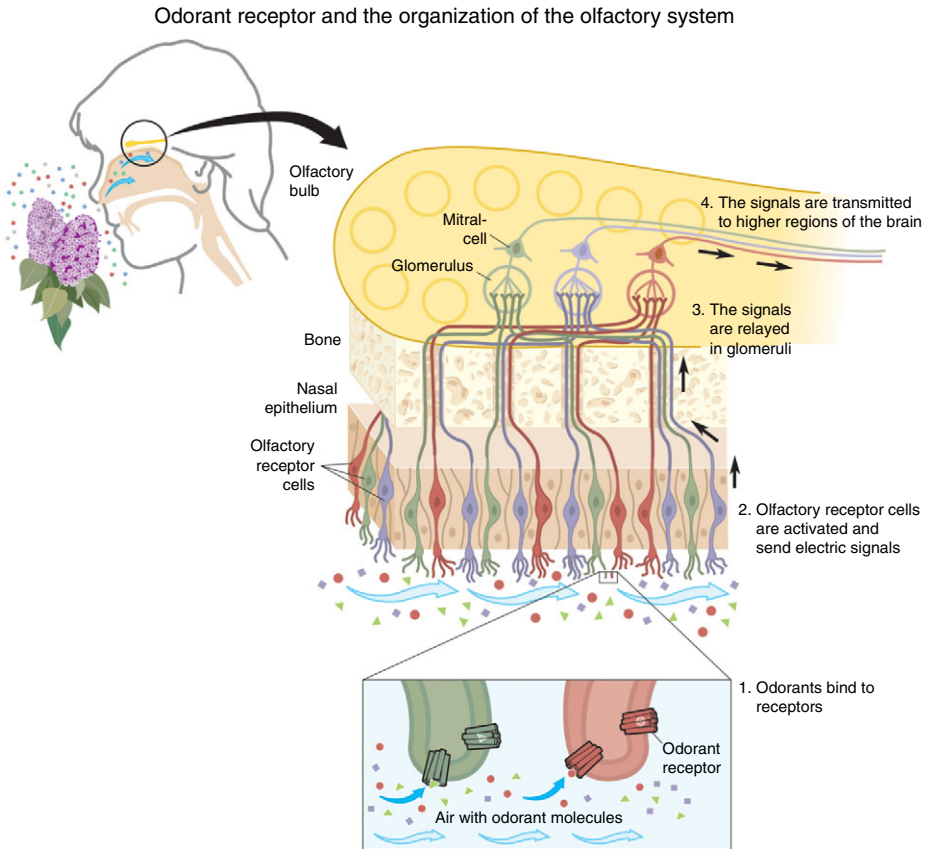

Odorant receptor and the organization of the olfactory system
Olfactory bulb
Mitral-cell
Glomerulus
4. The signals are transmitted to higher regions of the brain
3. The signals are relayed in glomeruli
Bone
Nasal epithelium
Olfactory receptor cells
2. Olfactory receptor cells are activated and send electric signals
1. Odorants bind to receptors
Odorant receptor
Air with odorant molecules

FIGURE 23.9 Schematic of the mammalian olfactory system. Olfactory receptor neurons are bipolar cells within the olfactory epithelium. Ciliated endings contain odorant receptors that detect odors. Axons project to glomeruli in the olfactory bulb. Mitral/tufted cells send this information to higher brain centers. *From the 2004 Nobel Prize in Physiology or Medicine, press release.*

problem of how diverse odors are recognized and distinguished, as discussed below.

G-Protein-Coupled Receptors Are the Initial Site of Odor Transduction in Mammals

The 2004 Nobel Prize in Medicine was awarded to Dr. Richard Axel and Dr. Linda Buck for their seminal discovery of a large family of mammalian olfactory receptors. Their search for olfactory receptors was based on three assumptions: first, that olfactory receptors would be G-protein-coupled receptors (GPCRs) similar to visual receptors; second, that they would be exclusively expressed in nasal epithelium; and third, that there would be a large number of receptors, given the large number of odors that mammals smell. This led Axel and Buck to search for novel GPCRs in nasal tissue and to the discovery of the odorant receptor gene (OR) family (Buck et al., 1991).

There are approximately 1,300 olfactory genes in mouse (5% of their genome), 500 in humans (2% of the genome), and 100 in zebrafish and catfish. They are members of the class A family of GPCRs, which includes opsin and β-adrenergic receptors. These receptors contain a short extracellular amino terminus, seven membrane-spanning (TM) domains, and an intracellular carboxyl terminus. Hypervariable regions in TM 3, 4, and 5 likely form the ligand-binding pockets, based on the ligand binding domains of other GPCRs such as the β-adrenergic receptor. Individual human olfactory receptor genes share between approximately 50 to 95% sequence identity. Large gene families comprising olfactory receptors have also been identified in nematodes (see Box 23.3) and insects (see below), although they bear little sequence similarity to mammalian ORs.

Olfactory receptor genes are organized in the genome in large linked arrays on several chromosomes. The size of the arrays ranges from 6 to 138 in humans and up to 15 in *C. elegans*. The *Drosophila* genome, with fewer olfactory receptors, contains small arrays of 2 to 3 genes, as well as many singly distributed genes. The large arrays in mammals and nematodes suggest that odor receptor gene families are rapidly evolving by duplication and expansion.

The identification of odor receptors provided fundamental insight into the molecular basis of odor perception. The large number of receptors dedicated to binding odor molecules provides an elegant solution to the problem of how animals detect a large number of smells. In addition, identifying receptors opened the door to examine interactions of different receptors with different odors, the organization of olfactory receptors in olfactory neurons, and the representations of olfactory receptors in the olfactory bulb.

BOX 23.3

SENSING THOUSANDS OF CHEMICALS WITH A HANDFUL OF NEURONS

The nemode worm *Caenorhabditis elegans* lives in the soil, lacks a visual system, and relies largely on chemical detection to survive in the environment. This small animal has only 302 neurons, but displays complex olfactory behaviors, making it a model system to study chemosensory processing. Thirty-two sensory neurons mediate chemical detection and contain members of a large chemoreceptor gene family. These receptors are G-protein-coupled receptors, but are not related to mammalian olfactory receptors. There are approximately 1,500 chemoreceptor genes in *C. elegans*, comprising 7% of all its genes. Thus, of all animals whose genomes have been sequenced, *C. elegans* dedicates the most genes, and the highest fraction of its genes, to chemical sensing.

Many receptors are found in the same chemosensory cell, as might be expected given the large number of receptors and the small number of sensory cells. This organization differs from the olfactory systems of mammals and flies but resembles mammalian and fly gustatory systems. Parceling many receptors into a single cell allows these animals to detect many chemical compounds despite the limited number of chemosensory neurons.

C. elegans shows sophisticated behavioral responses to chemical cues, including attraction and avoidance of different volatiles cues, adaptation to specific odors, as well as learned associations. For example, worms can learn to avoid pathogenic bacteria based on their smell after one exposure to the bacteria (Zhang et al., 2005). This is similar to conditioned taste aversion in mammals where one encounter with a nausea-producing food leads to aversion to the smell and taste of the food.

Kristin Scott

Evidence that ORs Detect Odors

The large number of ORs and the large number of odor molecules, as well as technical difficulties in expressing ORs in heterologous cells, has limited the ability to determine the ligands that different ORs recognize. In the first demonstration that ORs detect volatile cues, the "I7" odor receptor was overexpressed in most olfactory neurons of the rat. The electro-olfactogram (EOG) response of the epithelium to a battery of odors was examined. Among some 80 compounds tested, only longer-chain aldehydes gave increased responses over controls, with octyl aldehyde giving the peak response, and lesser responses for flanking long-chain aldehydes. This experiment demonstrated that a single OR detects a small subset of odors (Zhao et al., 1998).

More recently, advances in the ability to express receptors in heterologous cells have resulted in the identification of receptor-ligand pairs for several different ORs (Saito et al., 2009). Some ORs bind very few ligands (specialists), whereas others are broadly tuned (generalists). These receptor-ligand studies have shown that one OR will bind more than one odor and one odor will activate more than one OR. Thus, a single odor will activate multiple receptors and different odors will activate different combinations of receptors.

One Olfactory Receptor Gene Is Expressed in Each Olfactory Neuron

Studies of olfactory receptor expression in mammals revealed that each olfactory neuron contains only one member of the OR gene family. OR gene expression was determined by *in situ* hybridization experiments, showing that each OR is found in approximately 0.1% of olfactory neurons. In addition, isolating OR receptor sequences from single olfactory neurons demonstrated that each neuron expresses mRNA for only one OR. Finally, several pairs of ORs were examined to determine if two receptors were found in the same cell, but none of the ORs examined were co-expressed. Instead, receptors label nonoverlapping cell populations.

Not only is one receptor expressed per cell, but only one allele of a receptor is expressed per cell. An olfactory neuron expresses either the maternal or paternal allele but not both. The consequence of allelic exclusion is that each olfactory neuron contains only genetically identical receptors. Small changes in the nucleotide sequence of an allele might alter its ligand-binding properties, such that the maternal and paternal alleles recognize different odors. Segregating these alleles into different cells might expand the repertoire of odors that different olfactory neurons detect.

The general rule that each olfactory neuron expresses only one olfactory receptor gene has important implications for olfactory coding in the periphery. This organization dictates that the neural activity of a sensory neuron reports the activity of a single receptor and the activation of different olfactory receptors produces responses in different cells. Parcelling receptors into different cells allows the animal to distinguish which receptor has been activated by which cells have been activated.

Olfactory Signal Transduction Utilizes Adenylate Cyclase and a Cyclic Nucleotide Gated Ion Channel

The olfactory signal transduction cascade is similar to other G-protein-coupled signaling cascades (Fig. 23.10). Ligand-binding to an odorant receptor induces a conformational change that results in activation of a heterotrimeric G protein. The Gα subunit is an olfactory-specific G_s protein (G_{olf}) that activates an adenylate cyclase type III (AC3). AC3 is an enzyme that catalyzes the conversion of ATP to 3′,5′-cyclic AMP (cAMP). A cyclic nucleotide-gated (CNG) channel is activated in response to increases in cAMP, leading to cell depolarization, the generation of action potentials, and neurotransmitter release.

Mice lacking the G_{olf}, the AC3, and the CNG channel have been generated and the phenotypes of these mutants highlight the critical roles of olfactory signaling. Olfactory neurons from the AC3 mutants and the CNG channel mutants do not respond to odors. Mice lacking the G_{olf}, AC3, or CNG channel generally die within 1 to 2 days of birth without milk in their stomach, unable to nurse. This reduced survival rate is consistent with the idea that olfactory cues are necessary for newborn pups to suckle. These studies demonstrate that the G_{olf}, the AC3, and the CNG channel are essential components of the olfactory signal transduction cascade.

Axons of Olfactory Neurons Project to the Olfactory Bulb

Olfactory neurons in the nasal epithelium send axons directly to the mammalian brain, to paired structures in the most rostral region called the olfactory bulbs. The olfactory bulb is the primary relay station for smell. It is comprised of multiple cellular layers. Most peripheral is the glomerular layer where the axons of olfactory neurons reside, with deeper layers containing cells that synapse with olfactory neurons.

Axons of olfactory neurons terminate in the rounded regions of neuropil termed glomeruli. Anatomical studies from the time of Ramon y Cajal have shown that the glomerulus is an anatomical unit for the convergence of axons from many ORNs. In rats, the 15 million olfactory neurons converge onto 1,500 glomeruli, giving an average overall convergence of some 10,000:1. *In situ* hybridization and gene-targeting methods showed that ORNs expressing the same odorant receptor protein converge onto one or a few glomeruli in the olfactory bulb (Mombaerts et al., 1996) (Fig. 23.11). Neurons with different ORs send axons to different glomeruli. Activation of different olfactory receptors in the periphery leads to the activation of different glomeruli in the olfactory bulb. This suggests that the animal knows what it is smelling by the combination of glomeruli that are activated.

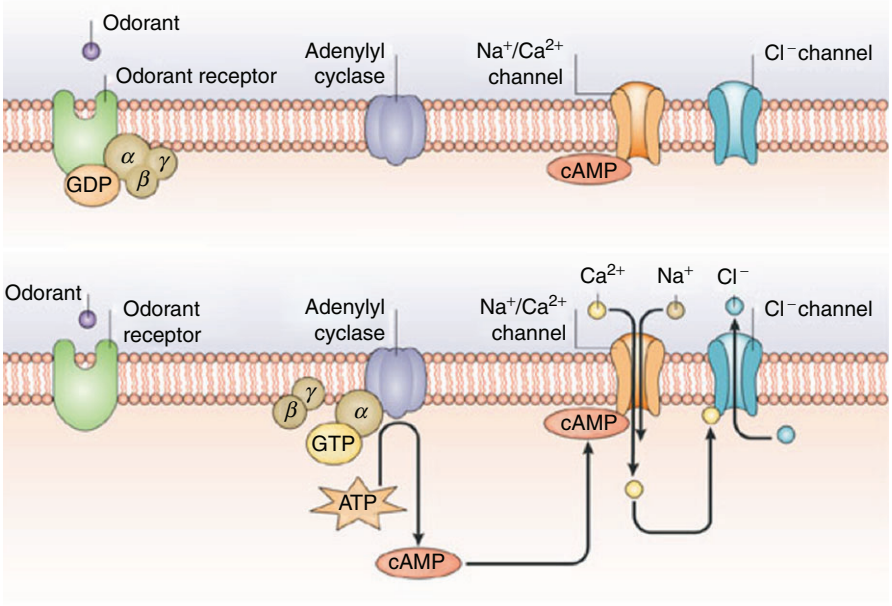


FIGURE 23.10 Sensory transduction of odor molecules involves a cyclic AMP second messenger pathway. Activation of a receptor by odor molecules activates a GTP-binding protein ($\alpha\beta\gamma$) that activates an adenylate cyclase, which produces cyclic AMP (cAMP), and a cationic cyclic nucleotide-gated channel and a chloride channel. *Reprinted by permission from Macmillan Publishers LTD:* Nature Reviews Neuroscience 5*(4): 263–78, copyright 2004.*

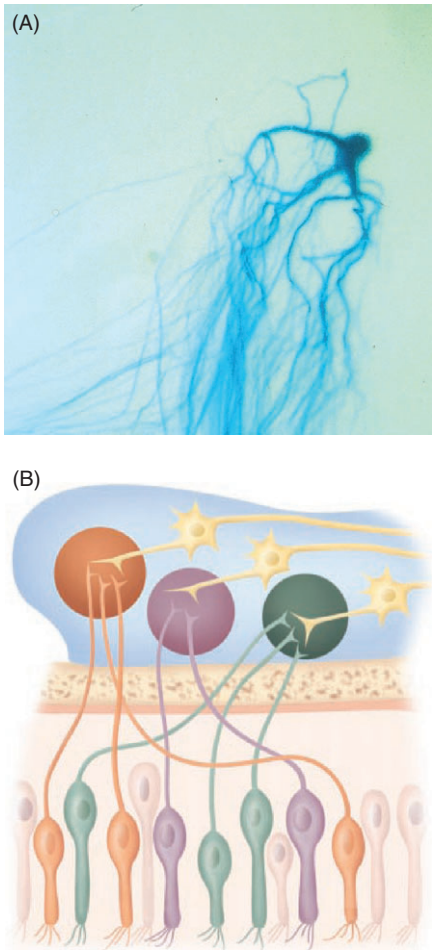


FIGURE 23.11 (A) Whole mount of the nose of a rat shows olfactory receptor axons converging on a single glomerulus. This subset of cells is stained for a lacZ reporter linked to mRNA for one type of olfactory receptor and the microtubule-associated protein tau. (B) Diagram of projections from the olfactory epithelium to the olfactory bulb. Neurons with the same receptor are scattered in the epithelium, but send axons that converge onto 1 or 2 glomeruli in the olfactory bulb. *From Mombaerts (1999). Reprinted with permission from AAAS.*

DIFFERENT ODORS ACTIVATE DIFFERENT COMBINATIONS OF GLOMERULI

The functional significance of glomeruli is seen when an animal is exposed to an odor and the activity patterns in the olfactory bulb are observed. The key observation is that different odors evoke different patterns of active glomeruli in the olfactory bulb. Thus, the focal glomerular activity elicited by the odor of clove is different from the pattern of glomeruli activated by coffee (Fig. 23.12). Some individual compounds are shared in coffee and clove and these activate the game glomeruli. Other compounds are not shared, leading to different overall patterns of activity. Each odor activates a different combination of glomeruli such that odors can be distinguished by their unique patterns of glomerular activation in the bulb (Xu et al., 2000). Thus, odors are thought to be represented by a combinatorial code in the olfactory bulb, where different combinations of glomeruli report different smells.

The Olfactory Bulb Is the Primary Relay for Smell

The olfactory bulb provides the first stage of synaptic processing of the sensory information in the olfactory pathway. The axons of olfactory neurons with the same odorant receptor converge onto 1 or 2 glomeruli in the olfactory bulb. Mitral/tufted cells are second-order neurons that synapse onto the olfactory axons and relay this information to higher brain centers. In addition to feedforward connections, there are also lateral connections. Periglomerular cells are inhibitory interneurons that make connections within and between glomeruli. Granule cells are interneurons that synapse onto mitral/tufted cells. The synaptic connections of different cell types in the olfactory bulb serve as the initial stage for olfactory information processing (Shepherd et al., 2007) (Fig. 23.13).

Within a glomerulus, olfactory axons make glutamatergic excitatory synapses onto the dendrites of the mitral/tufted neurons. Each mitral/tufted cell sends dendrites to one glomerulus, and thus relays the activity of one olfactory receptor to the brain. The dendrites of mitral/tufted cells also form reciprocal synapses with the periglomerular cell. The mitral/tufted dendrites make glutamatergic connections onto periglomerular cell dendrites, and dendrites of periglomerular cells form both GABAergic and dopaminergic synapses onto mitral/tufted cell dendrites. These connections are believed to mediate numerous types of interactions, including serial excitatory synapses (which spread excitation widely within a glomerulus), as well as recurrent and lateral inhibitory synaptic circuits (Urban et al., 2009).

In addition to intraglomerular processing, there is interglomerular processing. The axons of the periglomerular cells make synapses onto periglomerular cells in neighboring glomeruli and onto the dendritic shafts of neighboring mitral/tufted cells. One action of these synapses may be to inhibit periglomerular cells and mitral/tufted dendrites, thus providing contrast enhancement between neighboring glomeruli.

The second level of synaptic processing in the olfactory bulb occurs through inhibitory connections between granule cells and mitral cells. Reciprocal dendrodendritic synapses that mediate mitral/tufted-to-granule excitation and granule-to-mitral/tufted inhibition cause activated mitral/tufted cells to mediate feedback inhibition on themselves and lateral inhibition on their neighbors. Granule cells may participate in lateral inhibition onto a large population of neighboring cells that belong to neighboring glomerular units.

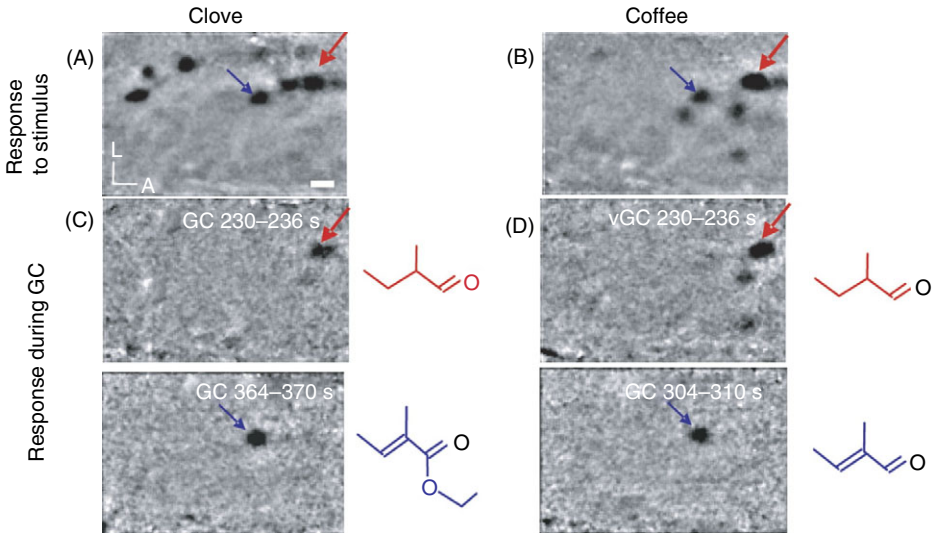


FIGURE 23.12 Odor maps. (A, B) Visualization of odor-elicited activity in glomeruli of the dorsal olfactory bulbs by intrinsic imaging for clove and coffee odors. Arrows mark glomeruli that are activated by compounds in both clove and coffee (C, D). Isolation of single compounds in coffee or clove odor, showing that the same compound is in coffee and clove and activates the same glomerulus. *From Lin et al. (2006).*

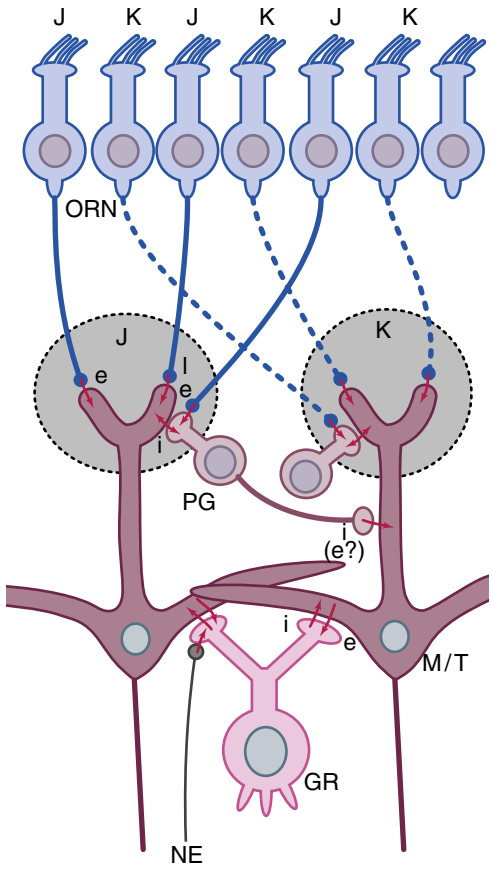


FIGURE 23.13 Diagram summarizing the synaptic organization of the glomerular layer. Olfactory neurons with the same receptor project to the same glomerulus. Mitral/tufted cells synapse onto a single glomerulus. Periglomerulur cells are inhibitory interneurons that synapse within and between glomeruli. Granule cells are inhibitory interneurons that synapse between mitral/tufted cells.

The current model is that the main function of the dendrodendritic synapses present at both the glomerular and granule cell levels is to mediate lateral inhibition and the role of this lateral inhibition is to enhance contrast between different odors.

Olfactory Bulb Output Goes Directly to Olfactory Cortex

The output of the olfactory bulb is carried by the axons of mitral cells and their smaller counterparts, the tufted cells. These axons project directly to olfactory cortex in the forebrain, the only sensory system to have this immediate access to the forebrain. Like cortical regions in other systems, the olfactory cortex is differentiated into several different areas (Fig. 23.14). The main area is the pyriform (sometimes called the prepyriform) cortex. This area receives input from mitral/tufted cells. It projects to the mediodorsal thalamus, which in turn projects to medial and lateral orbitofrontal areas of the neocortex. It is at this level that conscious perception of odors presumably takes place. Second is the olfactory tubercle, which receives input mainly from tufted cells. Third is the cortico-medial group of amygdalar nuclei, which receive specific input from the accessory olfactory bulb. Fourth is the lateral entorhinal area, which projects to the hippocampus. Last is the anterior olfactory nucleus, a sheet of cells just posterior to the olfactory bulb.

The *Drosophila* Olfactory System

The olfactory system of *Drosophila* shares many common principles with the mammalian olfactory system, although scaled down in size. The smell organs of the fly are its antennae and maxillary palp (Fig. 23.15). Olfactory neurons have dendrites that are exposed to the external

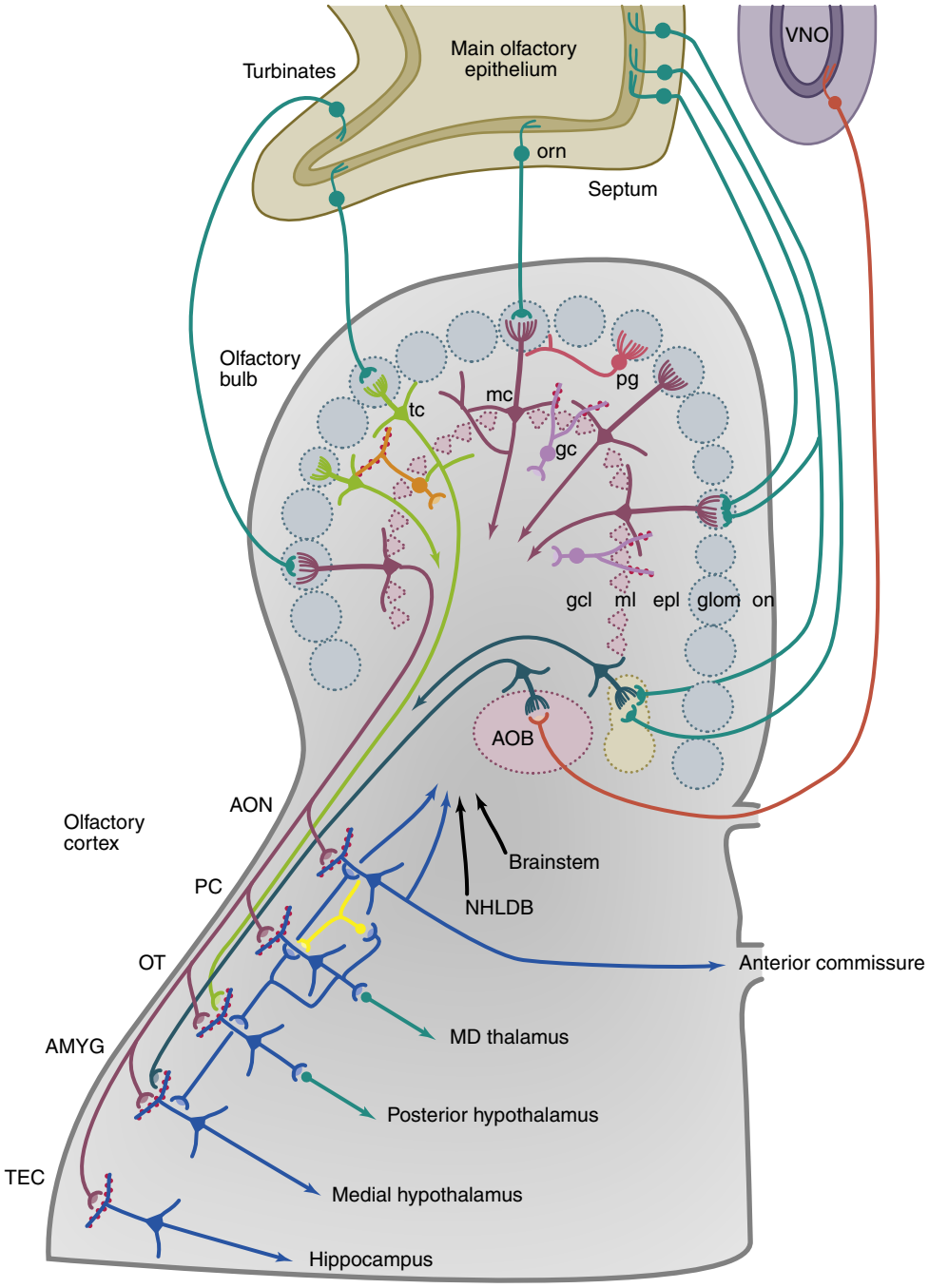


FIGURE 23.14 Summary of main projection pathways in the olfactory system. AON, anterior olfactory nucleus; PC, pyriform cortex; OT, olfactory tubercle; AMYG, amygdala; TEC, transitional entorhinal cortex; NHLDB, nucleus of horizontal limb of diagonal band; MD, mediodorsal.

environment for the detection of odors and axons that travel to the antennal lobe, the insect equivalent of the olfactory bulb. A family of approximately 60 *Drosophila* odorant receptors (dOR) detects volatile cues. Each receptor is expressed in a small subset of olfactory neurons, except for one receptor that is found in nearly all olfactory neurons. This receptor, dOR83b, is necessary to localize other odor receptors to the dendritic membrane and is essential for olfactory function (Larsson et al., 2004). Each olfactory neuron contains dOR83b as well as one other specific olfactory receptor. Neurons with the same specific receptor project to a single glomerulus in the antennal lobe. Thus the organization of the olfactory system in *Drosophila* is remarkably similar to the mammalian system in which neurons with different receptors project to different glomeruli.

Olfactory receptor neurons send axons to the antennal lobe. The second-order neurons are projection neurons. Each projection neuron sends dendrites to a single glomerulus and axons that branch to terminate in the protocerebrum (implicated in innate behavior) and mushroom bodies (involved in associative learning) of the fly brain. In addition, there are excitatory and inhibitory local interneurons in the antennal lobe that serve to shape the response of projection neurons. These interneurons act to amplify weak

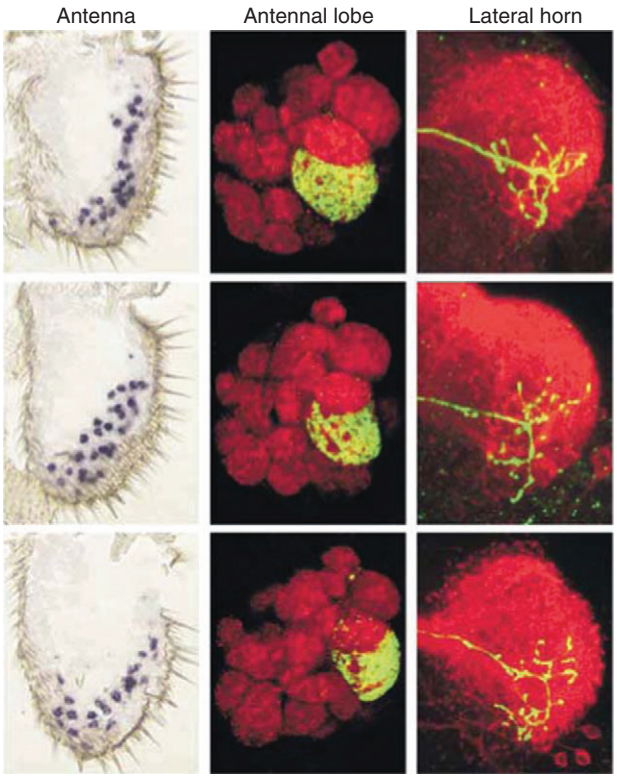


FIGURE 23.15 The fly olfactory system. A. The left panel shows a diagram of the fly head, highlighting olfactory regions. The antenna containing olfactory neurons is in light blue, and the first relay, the antenna lobes, is in dark blue. Neurons with the same receptor project to a single glomerulus. A single projection neuron innervates one glomerulus and sends axons to the protocerebrum. B. Olfactory neurons containing Or47a (blue) in the antenna. C. The projections of Or47a neurons (green) in the antennal lobe. D. The projection neuron that synapses onto Or47a neurons arborizes in a stereotyped fashion in the protocerebrum of the fly brain. *From Keller and Vosshall (2003).*

olfactory signals and dampen strong ones. Thus, they serve as a gain control mechanism to allow an animal to detect odors over a broad dynamic range (Wilson, 2011).

In general, although flies and mammals are separated by approximately 400 million years in evolution, the principles of olfaction are conserved in these organisms. In fly olfaction, as in mammalian olfaction, one receptor is expressed per cell, and neurons with the same receptor project to the same glomerulus. Comparative studies of the olfactory systems of different organisms highlight the general strategies that animals use for chemical recognition.

PHEROMONE DETECTION

The Accessory Olfactory System May Mediate Pheromone Detection

In addition to the main olfactory pathway, an accessory olfactory pathway exists for the detection of pheromones in many species. Pheromones are species-specific and gender

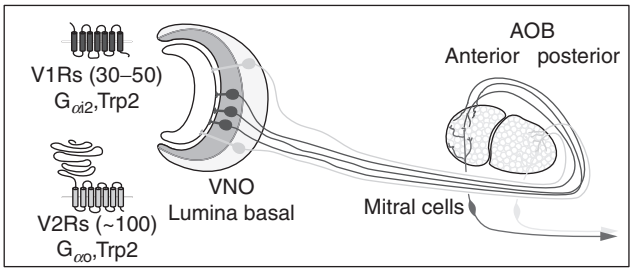


FIGURE 23.16 The accessory olfactory system of the mouse. The vomeronasal organ in the nasal cavity contains two layers. The apical layer expresses the V1R family of pheromone receptors, and the basal layer contains the V2R family. These neurons project to the accessory olfactory bulb (AOB) where they form several glomeruli. *From Dulac (2000).*

specific chemical cues emitted by an individual and detected by other members of the species that provide information about the individual's social, sexual, and reproductive status. In general, pheromones cause stereotyped and innate changes in conspecifics' behavior, either by direct, short-term effects on animal behavior or longer-term changes that alter the endocrine state. For example, one of the few vertebrate pheromones isolated is the sex pheromone of the red-sided garter snake. Female snakes produce this substance, which is sufficient to induce courting behavior in males. Another well-described pheromone is made by slave-maker ants. They emit a blend of chemicals that serve as an "alarm" signal, causing the slave-makers to attack other ants' nests, and the resident ants to flee.

The sensory structure for pheromone detection in mammals is the vomeronasal organ (VNO), also referred to as Jacobson's organ (Fig. 23.16). The VNO is a cigar-shaped cavity in the anterior nasal septum surrounded by bone. Openings in the anterior cavity contact either the oral cavity or the nasal cavity, depending on the species, for the detection of pheromonal cues. The sensory epithelium of the VNO is divided into anterior and posterior layers. Each layer contains sensory neurons that detect soluble and volatile cues. Different receptors and signaling molecules are found in the two layers, suggesting that they serve different functions.

Axons of sensory neurons travel from the VNO to the accessory olfactory bulb (AOB), which is located at the dorsal posterior surface of the main olfactory bulb (MOB) in rodents. Projection neurons then send axons to the bed nuclei of the accessory olfactory tract and stria terminalis, and the posteromedial cortical and medial nuclei of the amygdala. These projections do not overlap with projections from the main olfactory bulb, demonstrating that there is segregation of processing from the two systems.

Information Flow from the VNO to the AOB

The sensory epithelium of the VNO is divided into an apical and a basal layer. Neurons in the apical layer

express V1R receptors and those in the basal layer express V2R receptors (see following, Signal transduction in the VNO). In general, each sensory neuron expresses one receptor, similar to the olfactory system of mammals. This suggests that each sensory neuron detects a small subset of chemical cues.

Neurons from the apical layer project to the anterior region of the AOB and neurons from the basal layer project to the posterior AOB. Unlike the olfactory system, neurons with the same receptor do not project to one or a few glomeruli (Fig. 23.16). Instead, neurons with the same receptor converge onto multiple glomeruli (from 6 to 30) in spatially restricted domains, forming a complex pattern that is loosely conserved from animal to animal. Dendrites of mitral cells synapse onto several glomeruli in a complex fashion. Mitral cell dendrites have been reported to contact glomeruli from neurons with the same receptor as well as from neurons with different receptors. Thus unlike the olfactory system, a single mitral cell does not synapse onto a single glomerulus. The complex connectivity suggests that integration of different sensory inputs occurs in the AOB. One hypothesis is that this organization allows the animal to respond only to the relevant mixture of pheromone cues. For example, the simultaneous activity of different sensory receptors may be necessary to activate mitral cells, such that mitral cells respond to specific blends of pheromones. In this model, only the right mix of pheromones will trigger stereotyped innate behavior.

Signal Transduction in the VNO

Different signaling molecules reside in the apical and basal compartments of the VNO. In the apical layer, the V1R family of candidate pheromone receptors is expressed (Dulac et al., 1995). There are approximately 165 full-length V1R genes in mice, 106 in rats, and only 2 in humans. These G-protein-coupled receptors all contain a short amino-terminus extracellular domain and are distantly related to the T2R family of mammalian bitter receptors. Mice lacking 12 V1Rs do not detect a subset of candidate pheromones by electrophysiological studies of VNO field responses. These mice also show behavioral defects in stereotyped behaviors thought to be mediated by pheromones: for example, the female mice do not show maternal aggression (normally, nursing mothers fend off intruders). These findings argue that the V1R family detects pheromones and mediates behavior. Neurons that express V1Rs also contain the Gαi2 alpha subunit of heterotrimeric G proteins and the ion channel TRPC2, as shown by *in situ* hybridization and immunocytochemical staining. This suggests that these molecules transduce information from the activation of V1Rs to cell depolarization.

The V2R family of chemoreceptors is expressed in the basal layer of the VNO. This receptor family has a long amino-terminal extracellular domain and shares sequence similarity with the mammalian T1R family of taste receptors. There are approximately 60 intact V2R receptor genes in mice and rats. All genes are expressed exclusively in the VNO, suggesting that they mediate pheromone detection. The Gαo subunit of heterotrimeric G proteins is expressed in the basal layer along with V2Rs. In addition, the TRPC2 ion channel that is expressed in the apical layer is also found in the basal layer.

Mice Lacking TRPC2 Show Behavioral Defects

Because the TRPC2 ion channel is expressed in both the apical and basal layer of the VNO, it may act as an ion channel mediating detection of pheromones. To test this, mice were generated that lack the TRPC2 channel (Stowers et al., 2002). The VNO is not activated by candidate pheromones in these mice by electrophysiological recordings, demonstrating a deficit in VNO function. If the VNO is not functional, one prediction is that these mice will lack pheromone-driven social behaviors such as mating and aggression. Consistent with this, mice lacking TRPC2 do not show aggressive behavior. Surprisingly, male mice without TRPC2 indiscriminately mate both females and males. The simplest interpretation of these studies is that mating does not require the VNO; instead, the VNO is required to inhibit male-male courtship behavior and promote discrimination between the sexes.

Pheromone Detection in Humans?

Although evidence is accumulating that the VNO participates in pheromone-driven behavior in mice, the role of the VNO in human behavior is far less certain (Meredith, 2001). Anatomical studies of the vomeronasal organ argue that this structure exists but lacks sensory neurons and nerve bundles in the adult. The accessory olfactory bulb exists in the fetus but regresses and is not present in the adult. Genomic studies demonstrate the TRPC2 gene is a non-functional pseudogene in humans. Moreover, almost all V1R pheromone receptors are pseudogenes (115/117). Taken together, these studies strongly argue that the accessory olfactory system does not function in humans.

The loss of VNO in humans can be interpreted in two ways. One possibility is that humans do not detect pheromones and have evolved other strategies to ensure appropriate detection of mates and attackers. Alternatively, it is possible that humans detect odors through the main olfactory system that act as pheromones to elicit innate, stereotypical behaviors involved in reproduction and aggression.

Comparison of the Main Olfactory System and the Accessory Olfactory System

Although the accessory olfactory system has been classically thought to detect pheromones and the main olfactory system to detect a vast array of volatile cues, the distinction between the two noses is blurred both in terms of the compounds they detect and the behaviors they mediate. For example, electrophysiological studies show that sensory neurons in the VNO respond to volatile and nonvolatile compounds. Similarly, olfactory neurons depolarize in response to compounds classified as pheromones. Mice that have a defective VNO (lacking TRPC2) lack many pheromone-mediated behaviors, but not all. For example, pup suckling is a stereotyped behavior that is normal in TRPC2 mutants. Mice with a defective main olfactory epithelium (lacking the olfactory cyclic nucleotide-gated channel) do not mate, fight, or suckle, showing that the olfactory system is also required for innate stereotyped behavior. Thus, the best ways to distinguish the main and accessory olfactory systems are not by the ligands they recognize or the behaviors they mediate, but by their anatomical segregation in the periphery and central nervous system.

Summary

The olfactory system enables animals to detect and discriminate between thousands of different odors. The enormous number of olfactory receptors enables the ability to detect a vast array of odors. Each olfactory receptor recognizes a subset of chemical cues, with one odor activating more than one receptor and one receptor recognizing more than one odor. In the periphery, each olfactory neuron expresses only one olfactory receptor. Neurons with the same OR are randomly distributed in the nasal epithelium but they all project to the same glomerulus in the olfactory bulb. Thus, the activation of different olfactory receptors leads to the activation of different glomeruli in the central nervous system. Different combinations of activated glomeruli represent different smells. In *Drosophila* as well as mammals, one receptor is expressed per cell and neurons with the same receptor project to the same glomerulus, demonstrating that the logic of olfaction has been maintained through evolutionary time. In addition to the main olfactory pathway in vertebrates, a parallel accessory pathway exists for transmitting signals from less volatile odorous compounds called pheromones.

Higher order processing begins in the olfactory bulb, where mitral/tufted cells synapse onto olfactory neurons and transmit this information to the olfactory cortex. Periglomerular cells and granule cells provide inhibitory connections in the bulb that shape olfactory responses. Information is then relayed to five different brain regions where it is ultimately translated into different odor percepts and behavior.

References

Adler, E., Hoon, M. A., Mueller, K. L., Chandrashekar, J., Ryba, N. J. P., & Zuker, C. S. (2000). A novel family of mammalian taste receptors. *Cell, 100,* 693–702.

Amrein, H., & Bray, S. (2003). Bitter-sweet solution in taste transduction. *Cell, 112,* 283–284.

Buck, L., & Axel, R. (1991). A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. *Cell, 65,* 175–187.

Chandrashekar, J., Kuhn, C., Oka, Y., Yarmolinsky, D. A., Hummler, E., Ryba, N. J., & Zuker, C. S. (2010). The cells and peripheral representation of sodium taste in mice. *Nature, 464,* 297–301.

Dulac, C., & Axel, R. (1995). A novel family of genes encoding putative pheromone receptors in mammals. *Cell, 83,* 195–206.

Hoon, M. A., Adler, E., Lindemeier, J., Battey, J. F., Ryba, N. J., & Zuker, C. S. (1999). Putative mammalian taste receptors: A class of taste-specific GPCRs with distinct topographic selectivity. *Cell, 96,* 541–551.

Huang, A. L., Chen, X., Hoon, M. A., Chandrashekar, J., Guo, W., Trankner, D., Zuker, C. S., et al. (2006). The cells and logic for mammalian sour taste detection. *Nature, 442,* 934–938.

Keller, A., & Vosshall, L. B. (2003). Decoding olfaction in Drosophila. *Current Opinion in Neurobiology, 13,* 103–110.

Kim, U. K., Jorgenson, E., Coon, H., Leppert, M., Risch, N., & Drayna, D. (2003). Positional cloning of the human quantitative trait locus underlying taste sensitivity to phenylthiocarbamide. *Science, 299,* 1221–1225.

Larsson, M. C., Domingos, A. I., Jones, W. D. Chiappe, M. E., Amrein, H., & Vosshall, L. B. (2004). Or83b encodes a broadly expressed odorant receptor essential for Drosophila olfaction. *Neuron, 43,* 703–714.

Lin, da Y., Shea, S. D., & Katz, L. C. (2006). Representation of natural stimuli in the rodent main olfactory bulb. *Neuron, 50,* 937–949.

Meredith, M. (2001). Human vomeronasal organ function: A critical review of best and worst cases. *Chemical Senses, 26,* 433–445.

Mombaerts, P. (1999). Seven-transmembrane proteins as odorant and chemosensory receptors. *Science, 286,* 707–711.

Mombaerts, P., Wang, F., Dulac, C., Chao, S. K., Nemes, A., Mendelsohn, M., Axel, R., et al. (1996). Visualizing an olfactory sensory map. *Cell, 87,* 675–686.

Mueller, K. L., Hoon, M. A., Erlenbach, I., Chandrashekar, J., Zuker, C. S., & Ryba, N. J. (2005). The receptors and coding logic for bitter taste. *Nature, 434,* 225–229.

Nelson, G., Chandrashekar, J., Hoon, M. A., Feng, L., Zhao, G., Ryba, N. J., & Zuker, C. S. (2002). An amino-acid taste receptor. *Nature, 416,* 199–202.

Nelson, G., Hoon, M. A., Chandrashekar, J., Zhang, Y., Ryba, N. J., & Zuker, C. S. (2001). Mammalian sweet taste receptors. *Cell, 106,* 381–390.

Saito, H., Chi, Q., Zhuang, H., Matsunami, H., & Mainland, J. (2009) Odor coding by a mammalian receptor repertoire. *Science Signaling, 60,* ra9.

Shepherd, G. M., Chen, W. R., Willhite, D., Migliore, M., & Greer, C. A. (2007). The olfactory granule cell: From classical enigma to central role in olfactory processing. *Brain Research Reviews, 55,* 373–382.

Stowers, L., Holy, T. E., Meister, M., Dulac, C., & Koentges, G. (2002). Loss of sex discrimination and male-male aggression in mice deficient for TRP2. *Science, 295,* 1493–1500.

Urban, N. N., & Arevian, A. C. (2009). Computing with dendrodendritic synapses in the olfactory bulb. *Annals of the New York Academy of Sciences, 1170,* 264–269.

Wilson, R. I. (2011). Understanding the functional consequences of synaptic specialization: insight from the Drosophila antennal lobe. *Current Opinion in Neurobiology, 21,* 254–260.

Xu, F. Q., Greer, C. A., & Shepherd, G. M. (2000). Odor maps in the olfactory bulb. *The Journal of Comparative Neurology, 422,* 489–495.

Zhang, Y., Lu, H., & Bargmann, C. I. (2005). Pathogenic bacteria induce aversive olfactory learning in Caenorhabditis elegans. *Nature, 438*, 179–184.

Zhao, H., Ivic, L., Otaki, J. M., Hashimoto, M., Mikoshiba, K., & Firestein, S. (1998). Functional expression of a mammalian odorant receptor. *Science, 279*, 237–242.

Suggested Readings

Axel, R. (1995). The molecular logic of smell. *Scientific American, 273*, 154–159.

Bargmann, C. I. (2006). Comparative chemosensation from receptors to ecology. *Nature, 444*, 295–301.

Buck, L. B. (2000). The molecular architecture of odor and pheromone sensing in mammals. *Cell, 100*, 611–618.

Chandrashekar, J., Hoon, M. A., Ryba, N. J., & Zuker, C. S. (2006). The receptors and cells for mammalian taste. *Nature, 444*, 288–294.

DeMaria, S., & Ngai, J. (2010). The cell biology of smell. *Journal of Cell Biology, 191*, 443–452.

Dulac, C. (2000). Sensory coding of pheromone signals in mammals. *Current Opinion in Neurobiology, 10*, 511–518.

Dulac, C., & Torello, A. T. (2003). Molecular detection of pheromone signals in mammals: From genes to behaviour. *Nature Reviews Neuroscience, 4*, 551–562.

Hildebrand, J. G., & Shepherd, G. M. (1997). Molecular mechanisms of olfactory discrimination: Converging evidence for common principles across phyla. *Annual Review of Neuroscience, 20*, 595–631.

Mombaerts, P. (2004). Genes and ligands for odorant, vomeronasal and taste receptors. *Nature Reviews Neuroscience, 5*, 263–278.

Mori, K., & Yoshihara, Y. (1995). Molecular recognition and olfactory processing in the mammalian olfactory system. *Progress in Neurobiology, 45*, 585–619.

Mori, K., & Sakano, H. (2011). How is the olfactory map formed and interpreted in the mammalian brain? *Annual Review of Neuroscience, 34*, 467–499.

Scott, K. (2005). Taste recognition: Food for thought. *Neuron, 48*, 455–464.

Shepherd, G. M. (1994). Discrimination of molecular signals by the olfactory receptor neuron. *Neuron, 13*, 771–790.

Touhara, K., & Vosshall, L. B. (2009). Sensing odorants and pheromones with chemosensory receptors. *Annual Review of Physiology, 71*, 307–332.

Kristin Scott

CHAPTER

24

The Somatosensory System

The somatosensory system has many responsibilities. Beyond the obvious need to bring to consciousness the events that occur along the skin surface, this system provides an organism with knowledge of where it is in space, and it provides the motor system with feedback to control and coordinate action. Those are functions so diverse that minute variations on a common theme cannot accomplish them all. That is, the many demands on the somatosensory system cannot be met by subtle changes in gene expression or pattern of synapses or location of cell bodies. As a result, single neurons, large ensembles and groups of interconnected regions in the somatosensory system use different tactics and strategies to achieve the goals of perception, homeostasis and sensory guidance of movement.

PERIPHERAL MECHANISMS OF SOMATIC SENSATION

All Somatic Sensation Begins with Receptors and Ganglion Cells

Neurons of the dorsal root ganglia (DRG) and trigeminal ganglia are the only routes by which mammals receive information from the periphery of the body and of the face. These are classic pseudo-unipolar cells. Each has a single axon that divides close to the cell body and sends separate branches to the periphery and into the central nervous system (Fig. 24.1). Under normal circumstances neurons in somatosensory ganglia receive no synapses. Although that situation can change with damage to the peripheral branches of ganglion cells, the normal lack of synapses means the cell body of a somatic sensory ganglion cell is a factory for all that must be made to keep its two axon branches functioning properly. Its major function—to keep a long axon healthy enough to avoid conduction failure—it does very well, as the incidence for conduction failures in a ganglion cell axon is near zero. Even where failures are reported, as in the small ganglion cell bodies of some vertebrate species, the effect appears deliberate and is tied to the difference in action potential shape seen in large and small cells.

Neuronal cell bodies in somatosensory ganglia vary in size (Fig. 24.2). Although the details of size vary among species according to differences in body size, all species examined to date include two populations, one with somata at least half again larger than the cell bodies of the other. The population with smaller somata outnumbers those with larger somata by at least 2-to-1.

Peripheral branches of somatosensory ganglion cells enter peripheral nerves. In a mixed nerve, with both sensory and motor axons, the ganglion cells provide more than a fair share of the large, myelinated axons and all of the thinly myelinated and unmyelinated axons. Conduction velocities vary because of these differences in axon diameter and degree of myelination. Classic studies divide the sensory axons in human peripheral nerve into the following groups with conduction velocities in parentheses: Group I or Aα (70–120 meters/second), Group II or Aβ (40–70 meters/second), Group III or Aδ (12–36 meters/second), and Group IV or C (0.5–2.0 meters/second). These groups correspond to functional classes that carry proprioceptive, mechanosensory, thermoreceptive or nociceptive information from muscles, tendons and skin to the spinal cord (Table 24.1).

What sets apart ganglion cells of the somatosensory system from ganglion cells of other systems is the requirement for double duty. Not only do they carry information into the CNS, but they also transduce all forms of somatosensory energy into a change in membrane potential. That means somatosensory ganglion cells are also receptor cells. To do both, ganglion cells insert two types of cation

© 2013 Elsevier Inc. All rights reserved.

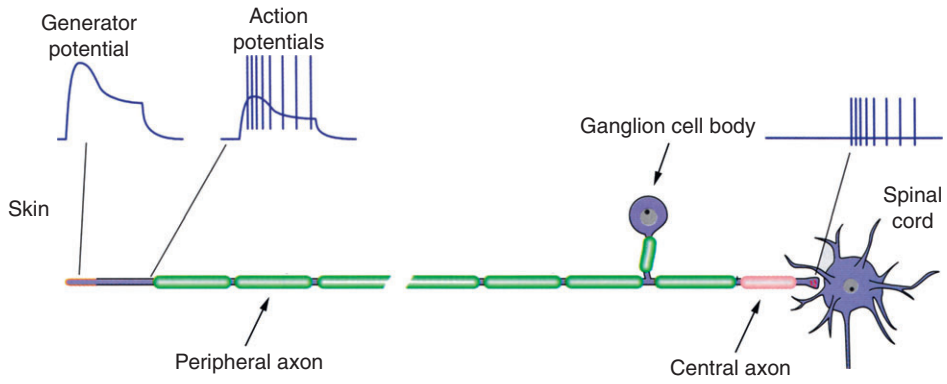


FIGURE 24.1 A dorsal root ganglion cell is a pseudo-unipolar neuron with an axon that divides at a T-junction into a peripheral branch and a central branch. At the tip of the peripheral branch are receptor proteins that, through opening of cation channels, produce a depolarization called a generator potential. With sufficient depolarization, voltage-gated Na^+ channels open to initiate action potentials. These action potentials are conducted down the axon and into the central branch that innervates second-order neurons in the spinal cord or in the medulla.

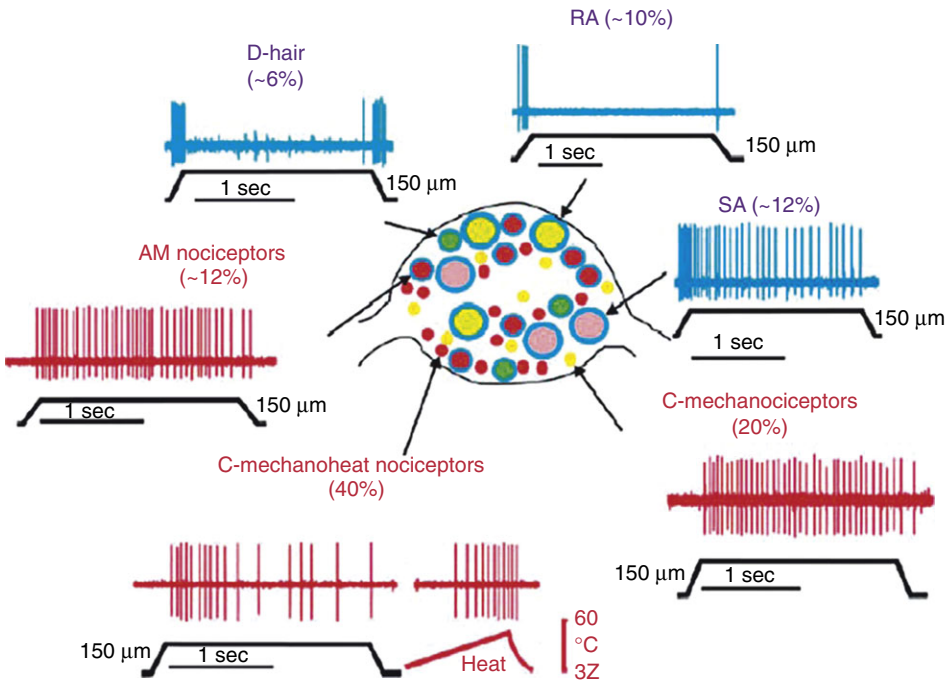


FIGURE 24.2 A somatosensory ganglion (center) is populated by a broad range of neurons that differ in size, gene expression and receptive field properties. Mechanoreceptors (in blue) differ in how they respond to a sustained stimulus. Rapidly adapting (RA) afferents of the glaborous skin and D-hair receptors or hair follicle receptors generate short bursts of action potentials at stimulus onset and offset. Slowly adapting (SA) afferents respond to a sustained indentation of skin with a prolonged series of action potentials. Nociceptors (in red) also vary in size and conduction velocity of their axons (Aδ vs. C) and in their response to noxious mechanical stimulation, such as a hard pinch. Some receptors respond specifically to this stimulus (AM nociceptors and C-mechanonociceptors) whereas others respond to a broad range of noxious stimuli (C-mechanoheat nociceptors). From Lewin and Moshourab (2004).

channels into the plasma membrane at the tips of their axons. One opens in response to a stimulus. Either it is physically tethered to peripheral tissue, and opens to an appropriate mechanical deformation, or it opens in response to changes in temperature, pH or the concentration of circulating factors that signal tissue damage. By way of these channels the tips of ganglion cell axons are depolarized to produce generator potentials. The second type of channel, voltage-gated Na^+ channels, turns generator potentials into action potentials that the axon conducts along its length and into the central nervous system. The best evidence suggests the voltage-gated channels are present at the first node of Ranvier in myelinated axons and at a similar distance from axon tip in unmyelinated axons (Fig. 24.1).

Many Types of Somatosensory Receptors Innervate Skin and Deep Tissues

The modern tradition breaks up the axons of sensory ganglia into 13 varieties of receptors. These include 4 types of proprioceptors, 3 types of nociceptors, 2 groups of thermoreceptors (one each for cooling and warming), and 4 types of mechanoreceptors. That is a decidedly conservative point of view. At a minimum it leaves unmentioned the class of D-hair receptor or hair follicle afferent and it recognizes only the major types of nociceptors. Molecular and functional distinctions among the nociceptors alone add enough additional types to make the list approach 20. The point here is not to split peripheral receptors into the greatest number of types but to

TABLE 24.1 Summary of Primary Afferent Fibers and Their Roles

Afferent type	SA1	RA	PC	SA2
Receptor	Merkel	Meissner	Pacinian	Ruffini
Location	Tip of epidermal sweat ridges	Dermal papillae (close to skin surface)	Dermis and deeper tissues	Dermis
Axon diameter	7–11 μm	6–12 μm	6–12 μm	6–12 μm
Conduction Velocity	40–65 m/s	35–70 m/s	35–70 m/s	35–70 m/s
Sensory function	Form and texture perception	Motion detection, grip control	Perception of distant events through transmitted vibrations, tool use	Tangential force, hand shape, motion direction
Effective stimulus	Edges, points, corners, curvature	Skin motion	Vibration	Skin stretch
Response to sustained indentation	Sustained with slow adaptation	None	None	Sustained with slow adaptation
Frequency range	0–100 Hz	1–300 Hz	5–1000 Hz	0–? Hz
Peak sensitivity	5 Hz	50 Hz	200 Hz	0.5 Hz
Threshold for rapid indentation or vibration (best)	8 μm	2 μm	0.01 μm	40 μm
Threshold (mean)	30 μm	6 μm	0.08 μm	300 μm
Receptive field area (measured with rapid 0.5 mm indentation)	9 mm^2	22 mm^2	Entire finger or hand	60 mm^2
Innervation density (finger pad)	100/cm^2	150/cm^2	20/cm^2	10/cm^2
Spatial acuity	0.5 mm	3 mm	10+ mm	7+ mm

emphasize the unusual nature of somatosensory transduction: the system responds to a broad range of stimuli because it has so many receptor types.

Mechanosensory Axons of Glabrous Skin Are Split into Four Types

Four types of mechanosensory axons are found in the hairless or glaborous skin and immediately adjacent deeper tissues of the human body (Fig. 24.3). Whereas a decade ago it would have seemed a radical thought, we now recognize the four receptors are responsible for specific functions. These include peripheral events that lead to perception of form and texture (Johnson, 2001), detection of object slip leading to adjustment of grip, sensory feedback necessary for the use of tools, perception of vibration and perception of hand shape and limb position (Table 24.1).

All mechanoreceptors conduct action potentials in the Aβ (type II) range, so no distinction can be made among them in that property. Rather it is in their response to the steady indentation of the skin that the difference among the four types becomes clear. Two types of afferents respond with a train of action potentials for as long as a blunt mechanical stimulus is applied to its receptive field. These are slowly adapting (SA) afferents, and they can be further divided into SAI and SAII types based on differences in spontaneous activity and in their sustained response to the applied stimuli (Fig. 24.2). Two other types of mechanoreceptor afferents are frequently called rapidly adapting (RA), or sometimes fast adapting or quickly adapting, because their response to a prolonged indentation of skin is a spike or two at onset of the stimulus, a single spike or two at offset and nothing in between. They, too, are split into two varieties, RA and PC, differing principally in their sensitivity and response to vibrating stimuli (see following). We should also point out some investigators object to the use of the term, adaptation, to describe this change in a mechanoreceptor axon's response because, as we will see, it is not the axon that adapts but the tissue around it. Nevertheless, the terms rapidly adapting and slowly adapting are illustrative and are in common usage, so will we use them here.

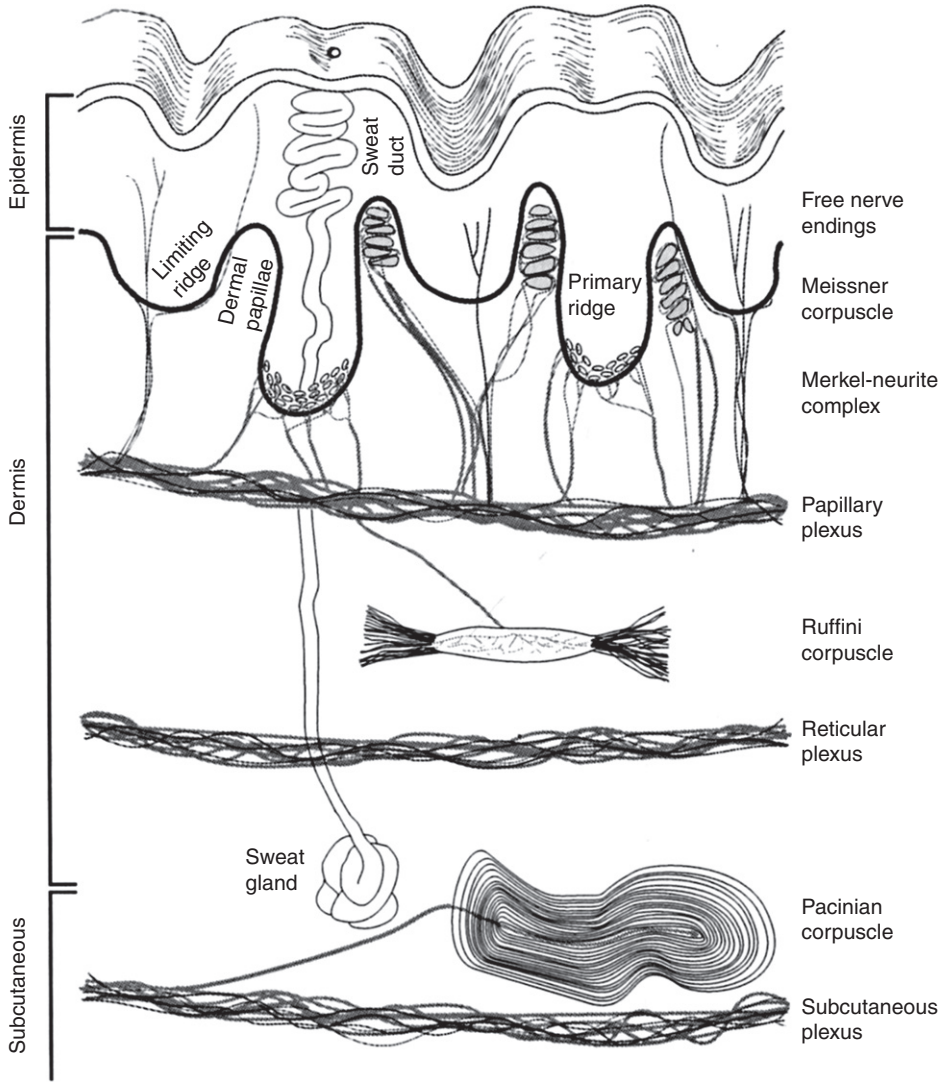


FIGURE 24.3 Peripheral receptors of the hairless (glaborous) skin are present in dermis, epidermis and subcutaneous tissue. Superficial receptors at the dermis-epidermis border include the free nerve endings of nociceptors and thermoreceptors, the rapidly adapting afferents associated with Meissner's corpuscle and the type I slowly adapting afferents that end as Merkel's disks. Deep receptors include a rapidly adapting receptor enclosed by a Pacinian corpuscle and a type II slowly adapting afferent in some species. Those SAII afferents are associated with Ruffini corpuscles in domestic cats but appear to have some other arrangement in most of the human hand. SAII afferents are missing from the skin of macaques and mice. From Johnson (2002).

Pacinian Corpuscles

The work of Iggo and Muir (1969) in the late 1960s established a consistent relationship among the four types of peripheral mechanoreceptor axons and the morphology of axon tips, or more properly of the axons' relationship with nonneural cells. For three of the axons this structure-function relationship has held up well. The clearest example is that of the PC afferent, the tip of which ends as part of a Pacinian corpuscle. Unmistakable in its onion skin-like structure, the Pacinian corpuscle filters out low frequency mechanical stimuli such as sustained pressure applied to the skin (Fig. 24.3).

PC afferents are exquisitely sensitive to vibration. When the frequency of vibration approaches 200 cycles/second (Hz), a PC afferent responds to skin indentations of no more than 10 nm. Yet a single, steady indentation of the skin surface produces only a couple of spikes from these axons. Careful dissection of the connective tissue that surrounds the PC axon shows that the axon itself adapts to only a limited degree. It stops responding because the fluid-filled capsule around it carries the energy of a continually applied probe away from the axon tip. That closes the cation channels responsible for mechanical transduction. By contrast, repeated application of a mechanical stimulus, such as a tuning fork that vibrates at 200 Hz, produces a series of discrete transduction events and a series of action potentials. We can say with great confidence, then, that PCs respond to high-frequency vibration at even the smallest magnitude. This extreme sensitivity to vibration turns the PC afferent into a detector of remote events: they respond to the minute vibrations of the skin produced whenever we use tools, such as pencils, screwdrivers, and keys.

Meissner's Corpuscles

Lower-frequency vibration, sometimes called flutter, produces a maximal response in RA afferents. As for PCs, the correlation between this type of response and

the structure of the afferent axon and its surrounding tissue is consistent. Each RA afferent ends as a stack of broad terminal disks within a Meissner's corpuscle composed of nonneural cells. Meissner's corpuscles exist in dermal pockets, as close to the epidermis as any dermal structure can be (Fig. 24.3), and their density is extraordinary, approaching 50/mm^2 in the index fingertip of a young adult. The result is an afferent very sensitive to even the slightest movement of skin, as happens when an object slides across the hand (Friedman et al., 2002). Yet every Meissner's corpuscle has at least two RA axons curled around it, and every RA axon innervates between 20 and 50 separate corpuscles. This convergence of two or more RA axons to a single corpuscle and divergence of a single RA axon to many corpuscles leads to large receptive fields (5 mm^2). So even though RA afferents are numerous and very close to the skin surface, their receptive field size and the filtering properties of the connective tissue capsule produce a response that is inappropriate for form and texture perception. RA afferents are responsible, instead, for the detection of objects slipping across the hand and fingers. They provide the sensory information that leads to the adjustment of grip force.

Merkel's Disks

Of the two slowly adapting afferents, SAIs end in a manner that prompts no arguments. A single SAI axon breaks into several branches that end in several closely packed dermal ridges (Fig. 24.3). Each branch ends in a series of axon terminals and each terminal is enfolded by an epidermal Merkel cell, all of which produces a surface elevation called a touch dome.

Merkel cells contact the tips SAI afferents by what looks to be a chemical synapse, and vesicles in the Merkel cell are filled with neurotransmitters. When a gene mutation deletes Merkel cells early in embryonic development—before ganglion cell axons reach the skin surface—SAI responses fail to develop; when deletion occurs later in development, SAI responses persist. These data indicate Merkel cells play a fundamental role in the development of SAI afferents but contribute little to the transduction of mechanical stimuli.

SAI Afferents Are Responsible for Form, Texture, and Curvature

SAI afferents are the source of peripheral information used by the CNS to perceive form, texture, and object curvature. Form perception has been carefully studied by a combination of psychophysics and neurophysiology because the dimensions of form are readily quantifiable. When the performance of individual receptors on the surface of a monkey's finger pad is compared with the performance of the monkey itself, only one type of afferent is found to be sensitive enough to account for an animal's ability to discriminate the form of a mechanical stimulus. That receptor is at the tip of SAI afferents ending in Merkel's disk. One example is seen in the response to the type of dots used in Braille (Fig. 24.4). Only the SAI afferents respond to the minute edges present in these embossed dots (6.0 mm high or greater) so accurately as to account for the ability a monkey or a human has in telling one pattern from the next.

In contrast to tactile form, texture has relatively few dimensions (rough-smooth, hard-soft, and sticky-slippery are the most prominent), but they are difficult to quantify. Nevertheless, a series of careful studies has documented that for variations in roughness, only the response of SAI afferents matches human perceptual ability. That is, the variation in firing rates among SAI follows precisely the perception human subjects have of surface roughness. Much the same can be said for the detection of surface hardness. Only the response of SAI afferents can account for human perception of how hard or soft is the surface of an object scanned by the fingertips. In summary, a combination of physiological and psychophysical studies leads one to conclude that SAIs provide the central somatosensory system with all the information it needs to detect the shape, hardness, and the roughness of objects pressed or scanned across the skin.

SAII Afferents

A second slowly adapting afferent in human skin, the SAII afferent, differs from an SAI in the greater size of its receptive field, the reduced sensitivity to simple indentation of the skin, and the greater sensitivity to skin stretch.

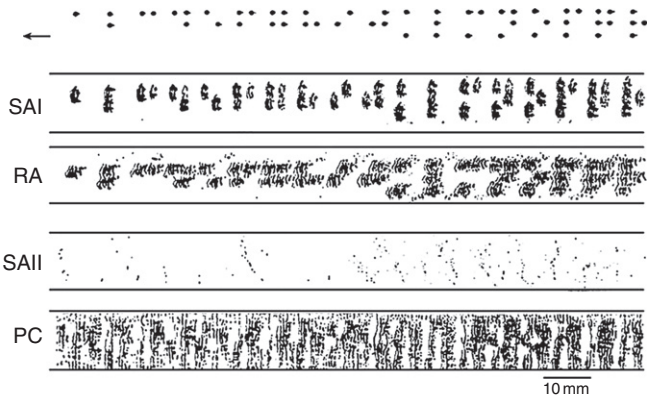


FIGURE 24.4 Responses of peripheral axons to a Braille pattern of dots scanned over the surface of a human fingertip at a rate of 60 mm/s with 200-μm shifts in position after each pass. Dots represent individual action potentials. Only the response of the SAI afferents (Merkel disk receptors) follows the Braille pattern faithfully, whereas RA afferents and Pacinians (PC) produce a response that distorts the input. SAIIs display little response to this stimulus. Adapted from Phillips *et al.* (1990).

One surprising feature of these afferents is their less-than-universal presence across the few mammals that have been studied. Direct recordings from the peripheral nerves of humans and domestic cats show these afferents to be a commonly encountered feature of both species, but they are rarely encountered in monkeys or mice. Just as perplexing is the poor correlation between structure and function with this receptor. SAII responses in cats arise from axons terminating in skin as Ruffini endings. These structures were first described in human skin at the turn of the twentieth century, but a very recent study found true Ruffini endings in human skin very rarely and only in the bed of fingernails (Pare et al., 2003). And so it appears SAII responses over most of the human hand arise from some other interaction of mechanoreceptor axon and connective tissue sheath. A conservative conclusion from these findings is that an arrangement of nonneural cells, perhaps a classic Ruffini corpuscle in cats but another configuration in humans, leads to the cardinal feature of SAII afferents—namely, a robust response to skin stretch. That configuration could vary from one species to another and from one area of skin surface to the next, but in the end, the nonneural tissue serves as a mechanical filter. After subtracting from the activity of a SAII afferent any response to simple deformation of the skin, the nonneural tissue leaves behind an unambiguous response to anything that stretches the skin.

Hair Follicle Afferents

In addition to receptor types found in glaborous skin, hairy skin is innervated by a separate receptor, called the D-hair receptor or hair follicle afferent (HFA). It is the most sensitive receptor in hairy skin. The HFA threshold is said to be one-tenth that of any other afferent in mouse skin, and displacement of a hair follicle by as little as 1 μm produces robust responses in this population of receptors. Single afferents innervate more than one hair follicle and as a result, the receptive field of an HFA is large (>10 mm^2 in mice). Unlike other mechanoreceptors, HFAs conduct action potentials in the Aδ range, which translates into a velocity of 20–25 m/s for humans but a much lower velocity in mice.

Receptor Density

A plot of SAI and SAII receptors along the human skin surface shows receptive field size and density vary by a factor of four in the short distance from the fingertip to the wrist. As is the case for other sensory systems, wherever receptor density declines, receptive field size ascends. This inverse variation in receptive size and density is reflected in the ability of a human subject to discriminate the two-dimensional shape of an object. Regions of high spatial acuity include the tips of the fingers and the face, whereas regions of low spatial acuity include the thigh and back.

Muscle Spindles and Golgi Tendon Organs Are Proprioceptors

The most prominent receptors fulfilling the function of sensing position and movement are muscle receptors and tendon receptors. These two have in common their sensitivity to stretch and the large diameter of the axons that carry the receptors' activity into the CNS. The manner in which they are arranged, however, makes all the difference in the world. Muscle receptors are arranged in parallel with the muscle fibers and as a result, these afferents respond when the muscle is stretched. As a common occurrence, muscle fibers stretch when load is added to them in the form of weight or resistance. The resulting stretch of the extrafusal or work muscle fibers produces a simultaneous stretch of the much smaller intrafusal muscle fibers. The intrafusal fibers have their own motor innervation and a sensory innervation from the largest diameter axons in a sensory nerve (Fig. 24.5). These sensory axons are called Ia afferents and they end in two configurations around the non-contractile portion of an intrafusal muscle fiber, where they signal the static and dynamic aspects of muscle stretch. A connective tissue capsule surrounds all of these—sensory axons, intrafusal muscle fibers, and motor axons—to form a muscle spindle. What the motor axons do to all this is simple to envision: by adjusting the contractile state of the intrafusal muscle fiber, they adjust the sensitivity of the muscle afferent. When the intrafusal fiber is contracted the spindle is at its most sensitive and when the intrafusal fiber is relaxed, the spindle is least sensitive.

Much as muscle receptors are sensitive to muscle stretch, tendon receptors (referred to as Golgi tendon organs) are sensitive to tendon stretch and provide information about muscle force. Because tendons are arranged in series with the muscles, so too are tendon receptors. What stretches the tendon is muscle contraction. In this regard, muscle spindles and Golgi tendon organs signal opposite trends. Spindles fire when a muscle is relatively inactive and stretched whereas tendon organs fire when the muscle is most active and contracted.

Proprioception Involves More Than Proprioceptors

By a definition of proprioception as an awareness of the position of the body and limbs, the two most prominent proprioceptors are necessary but not sufficient. More than just muscle and tendon receptors are at work to provide us with information about limb position. The related sense of kinesthesia, or awareness of position of moving body parts, is equally complex. Often used interchangeably for position sense, these two are percepts that require the contribution of receptors in muscle, skin and joints as well as a sense of muscle exertion. Studies in the

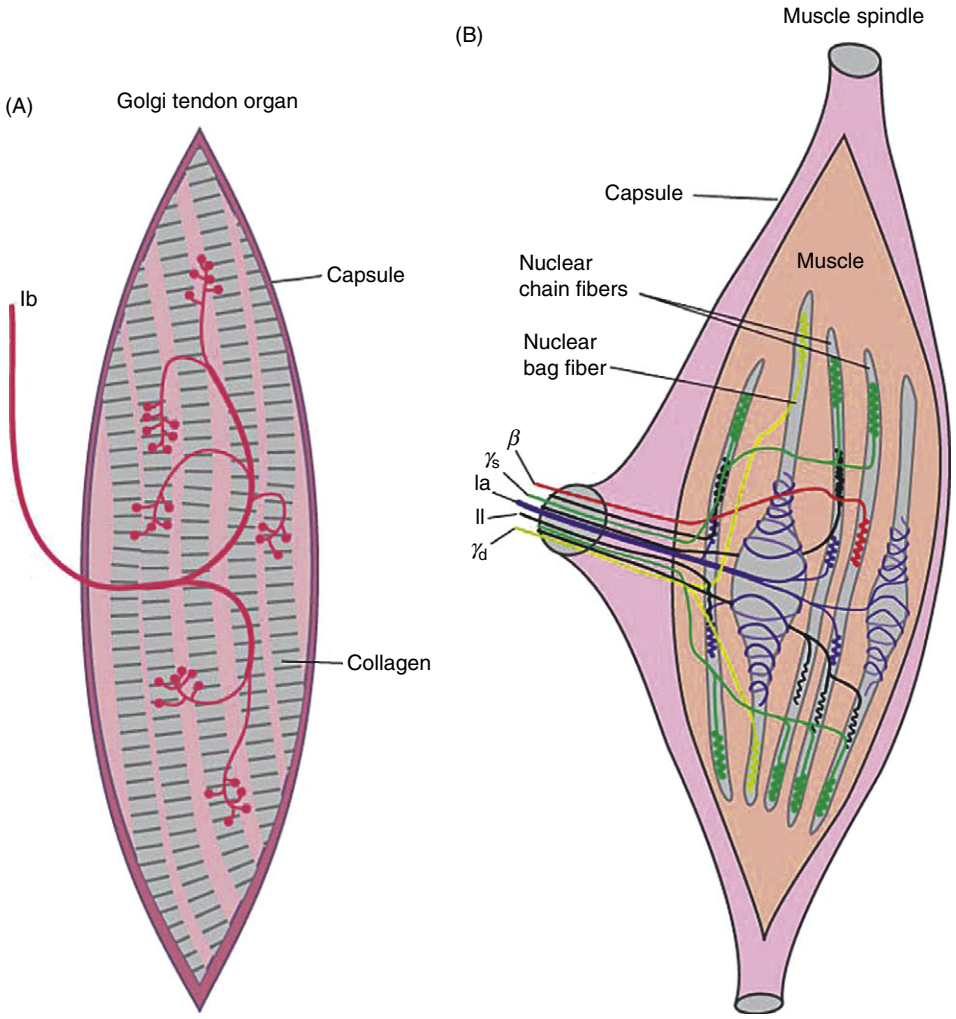


FIGURE 24.5 Proprioceptive afferents. (A) Golgi tendon organs and afferent termination along collagen fibers of the tendon capsule. These afferents respond when the entire capsule is stretched, usually by overvigorous contraction of the muscle. (B) Muscle spindle afferents (Ia and II) terminate on the noncontractile portions of intrafusal muscle fibers. They are arranged in parallel with working muscle fibers and respond to stretch of the entire muscle. Specialized motoneurons (γ) provide the motor innervation of the intrafusal muscle fibers and control the overall sensitivity of the muscle spindle.

1950s and 1960s focused on the contribution of joint afferents to a sense of limb position. Yet a series of findings in the 1970s made it clear that a sense of position survives joint removal; the joint afferents themselves respond only at the extreme limits of joint flexion. With no response in the usual, midrange of joint movement, these afferents could not signal position under most circumstances.

Studies of more recent vintage have dealt with muscle spindle afferents as a major source of the position signal. Since all muscles are organized as antagonistic pairs, contraction of one delivers a robust stimulus (muscle stretch) to afferents of the antagonistic muscle. From these sorts of inputs, limb position and movement would appear to require a simple neural computation. That position sense is significantly affected by stimulation, anesthesia and disengagement of spindle afferents adds weight to the argument that signaling limb position and movement requires these receptors. Perhaps the most convincing data are from studies of illusory movements produced by tendon vibration. This type of stimulus selectively activates spindle afferents and leads to the perception of limb movement when none has occurred. Illusory movement is muscle specific, so activation of arm flexors gives rise to the percept of an arm that has extended (a movement that normally produces activation of flexor muscle spindles). The illusion is so strong it gives rise to the Pinocchio effect: if arm flexor afferents are activated when a subject is touching his or her nose with an index finger, the nose itself is perceived to grow.

Only recently has proper attention been paid to the role of cutaneous receptors in position sense. SAII afferents, in particular, respond in a unique manner to any position a limb or digit might adopt. Particularly compelling is the role these skin stretch receptors play in signaling the position of fingers. For the hand and fingers, therefore, a conscious sense of position appears to arise from the cooperative activity of SAII afferents, muscle afferents and (at the extremes of movement) joint afferents.

NOCICEPTION

Nociceptors Respond to Noxious Stimuli

For most purposes, anything that has produced tissue damage or that threatens do so in the immediate future is defined as noxious; the type of axon that responds selectively to the noxious quality of a stimulus is a nociceptor. These are not pain receptors because nociception is not pain. Whereas nociception is a peripheral phenomenon that occurs when tissue has been or soon will be damaged, pain is a percept that requires regions of the CNS to become active.

All nociceptors end simply in skin and other peripheral tissues. These are usually described as free nerve endings because, unlike mechanoreceptive afferents, nociceptors do not end in a specialized capsule of nonneural cells. Nothing extraneuronal serves as a filter or buffer between the nociceptive axon tip and its immediate environment. The only thing that determines the response of a nociceptor, then, is the type of protein receptor and ion channel it inserts into its plasma membrane.

Nociceptors also differ from mechanoreceptors in how broadly the terminals of the peripheral axon branch as they reach target. Take the tip of the human index finger as an example. An individual SAI and SAII axon ends in a well-confined cluster of terminals over a distance as small as a few millimeters. Terminals of a single C fiber, by contrast, end over an area of more than a dozen millimeters. This is the first of several anatomical features along the nociceptive pathway that produce a much coarser spatial sense for pain than exists for mechanosensation. Only the simultaneous activation of mechanoreceptors as occurs with puncture wounds or damaging compression permits a person to accurately detect the location of a nociceptive stimulus.

The afferents that make up the nociceptive population can be subdivided into groups that are named by their axon conduction velocity (Aδ vs. C) and the response to noxious mechanical stimuli and noxious heat. Thus, an AM receptor conducts in the Aδ range and responds to intense mechanical stimuli, whereas a CMH receptor conducts in the C range and responds to both noxious mechanical energy and noxious heat. Other permutations of conduction velocity and response type are evident in the peripheral nerves of humans and other mammals, but these two are most common. They are frequently referred to as specific mechanical nociceptors and polymodal nociceptors.

First and Second Pain

Afferents responding specifically to a mechanical stimulus (AM receptors) carry the most rapid signals into the CNS, whereas those responding to the broad range of noxious stimuli (CMH receptors) conduct action potentials more slowly. These are the peripheral components to the two very different qualities of pain perceived by humans (Fig. 24.6). First pain or epicritic pain is rapidly perceived and carries with it much that is discriminative. A person can quickly and with some ease figure out what has happened and where it has happened when he or she drops a heavy object onto a toe or touches a hot stove surface with a couple of fingers on the right hand. First pain is informative and the peripheral component of it is the population of Aδ nociceptors. What follows later is second pain or protopathic pain. This is agonizing pain that carries much less information about location or source of energy. Second pain is punishing pain that serves to change the behavior of a person. Its peripheral component is the population of C nociceptors.

Two varieties of C nociceptors are found in mammalian skin. One variety is characterized by the presence of fluoride-resistant acid phosphatase (FRAP) in its cytoplasm and cell-surface glycoproteins recognized by the isolectin I-B4 and the monoclonal antibody LA4. Because none of these proteins is known to contribute to the physiological features of this C fiber type, their presence is currently a convenient feature that allows anyone studying them to recognize them. The situation is very different in the second type of C nociceptor. Present in its cytoplasm are two neuroactive peptides: calcitonin gene-related peptide (CGRP) and substance P. These play major roles in the function of the peptide-containing type of C nociceptor.

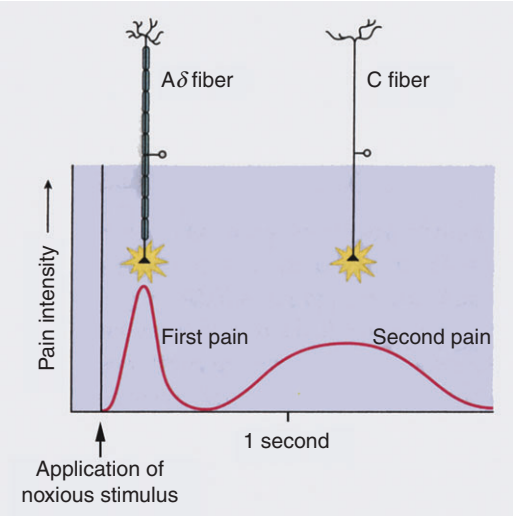


FIGURE 24.6 The two classes of nociceptors that conduct action potentials in the C and Ad ranges are peripheral components for two types of pain. First pain carried by Ad axons reaches consciousness rapidly and is discriminative. Both the location and the subjective intensity of the stimulus can be judged with relatively good precision in first pain. Second pain, in contrast, is much slower and is agonizing pain, with greatly reduced discriminative value.

The Axon Reflex

Release of neuropeptides from the second type of C nociceptor is responsible for the axon reflex. Injury to the skin surface is often well confined, as happens with a paper cut, for example. Yet within a short time of that injury, tissue surrounding the cut becomes reddened, in what is referred to as flair, and edema or swelling sets in as the tissue fills with fluid. Most importantly, the region surrounding a punctate wound becomes painful to touch even though it is outside the zone of direct damage.

Recall that an individual C fiber terminates over a wide area in skin, and so it is very likely that a punctate wound directly affects only a fraction of all that fiber's many branches. Action potentials generated at those directly affected branches invade all the peripheral branches, as well as the parent axon that conducts the signal to the CNS. At all the peripheral terminals, substance P and CGRP are released onto two principal targets: the smooth muscle surrounding peripheral blood vessels and histamine-rich mast cells. By causing the arterial smooth muscles to relax, the peptides increase the flow of blood into the damaged tissue and produce a flow of water and electrolytes out of capillaries and into extracellular space. This process is referred to as extravasation. Histamine released from mast cells leads to a pronounced inflammatory response. All of this is important for the infiltration of damaged tissue with cellular elements that will protect against infection and promote repair. Yet what is most obvious about the axon reflex is the much greater sensitivity of the tissue surrounding a wound to anything that might be noxious. This primary hyperalgesia is a direct result of the axon reflex. It occurs because the protein receptors inserted into nociceptive axons are sensitive to chemical changes produced by that reflex (Sandkiller, 2009).

Both the histamine released by mast cells and the edema resulting from extravasation affect the response of nociceptors. Histamine's effects are more selective, as only a subclass of the most slowly conducting C fibers insert histamine receptors into the membranes of their axon terminals. A considerable body of evidence suggests these are peripheral receptors whose activity leads to the unpleasant percept of itch. A parallel system for itch that operates independently of histamine release has been discovered recently but the population of afferents involved is not yet known.

Edema causes a general reduction in pH of extracellular fluid from 7.4 to below 6.0. A general feature of protein receptors inserted into nociceptive axons is their sensitivity to the concentration of H^+. Through this sensitivity to pH, activation of one branch of a C fiber leads to increased sensitivity of all its branches and all neighboring nociceptors to noxious stimulation. At least some of those neighboring nociceptors are likely to be silent nociceptors, so named because they are unresponsive to intense mechanical stimulation under normal circumstances. When activated by changes in pH or the presence of factors that signal the presence of tissue damage, the silent nociceptors become responsive to noxious stimuli.

Chemical Sensitivity of Nociceptors

Two of the most powerful pain-inducing chemicals that can be delivered to a human observer are natural products of tissue damage. They are lipids of the prostaglandin family and the peptide bradykinin. Prostaglandins are derivatives of the membrane fatty acid, arachidonic acid, which is itself a major component of the lipid bilayer. Damage to tissue and the resulting disruption of cell membranes releases arachidonic acid into extracellular fluid. The enzyme cyclo-oxygenase (COX) breaks down arachidonic acid to form prostaglandin. Inhibitors that target COX are referred to as nonsteroidal anti-inflammatory drugs (NSAIDS) and they include both aspirin and ibuprofen. Their popularity is a powerful indication of effectiveness in suppressing pain by interfering with the production of prostaglandins.

Nociceptor Proteins

Detection of noxious heat occurs by way of channels that are members of the transient receptor potential (TRP) family. Two members of the TRPV class—they are referred to as TRPV1 (Fig. 24.7) and TRPV2—pay particularly important roles in nociception (Tominaga et al., 1998). The V in the channel's name refers to its sensitivity to vanilloids, the active ingredients of chili peppers. TRPV1 is a nonspecific cation channel with approximately ten times more permeability to Na^+ than to Ca^{2+}. Like other members of the TRPV subfamily, TRPV1 is a molecular thermometer that opens in response to increased temperature. It has a threshold of 42^o C. TRPV2 channels, by contrast, open only when skin temperature reaches 50^o C. Although they are primarily detectors of noxious heat, nociceptors with TRPV1 channels are typically polymodal (they respond to other noxious stimuli), in part because of the TRPV1's properties and in part because these afferents express other receptor proteins. Because H^+ directly gates TRPV1 channels, any reduction in pH (the acidosis that accompanies tissue swelling) comes to depolarize these afferents. Other channels, the Acid Sensing Ion Channels (ASICs), also open in response to an increase in H^+ concentration. One of these, referred to as ASIC-3, is especially sensitive to the presence of lactic acid and, therefore, is a prime candidate for a receptor of ischemic pain in cardiac muscle.

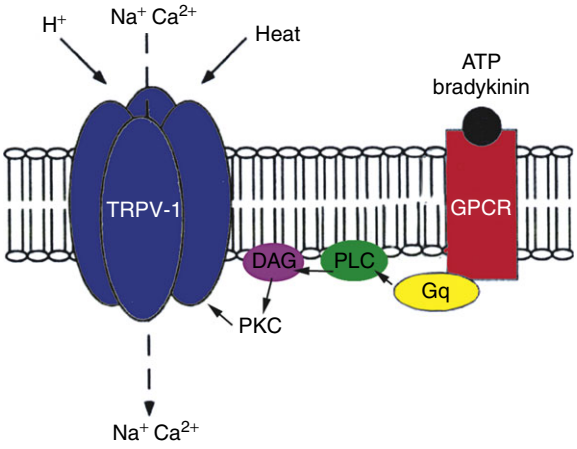


FIGURE 24.7 The receptor protein, TRPV1, provides a nociceptor with the ability to respond to many noxious stimuli. In addition to heat, TRPV1 is directly gated by a reduction in pH (afferent of H^+) produced in response to tissue swelling. Either or both opens a nonspecific cation channel that, through an influx of Na^+, depolarizes the nociceptor axon. Circulating agents that signal the presence of tissue damage (ATP and bradykinin) bind to a G-protein coupled receptor (GPCR). Through a series of steps PKCε is activated and TRPV1 subunits are phosphorylated, leading to a sensitization of the receptor. One result of this cascade is an opening of the cation channel at body temperature.

Two other proteins, both G-protein coupled receptors, broaden the response of polymodal nociceptors to include anything that produces tissue damage. One is a receptor for two molecules—bradykinin and ATP—released when the plasma membranes of cells are damaged. Their binding to plasma membrane receptors sets up an intracellular cascade that lowers the threshold temperature at which TRPV1 channels open. As that threshold descends to normal body temperature, TRPV1 channels open without any thermal stimulus at all and initiate action potentials in nociceptor axons. Finally, most polymodal nociceptors express GPCRs that bind prostaglandin and drive open an unusual population of Na channels. This, too, provides the afferents with a depolarizing stimulus when tissue is damaged.

One major outstanding question is the nature of the channel that opens in response to noxious (high-threshold) mechanical stimulation, such as a hard pinch. The same question exists for the low-threshold mechanically gated channel in SA and RA afferents, in proprioceptors, and in other cells, such as hair cells of the inner ear. Whereas common wisdom has long thought the somatosensory mechanoreceptors to be members of either the Epithelial Sodium Channel (ENaC) family or the TRP superfamily, very recent work suggests they are part of a novel group that has been dubbed Piezo. Caution is advisable, however, because the search for a mechanically gated channel has had its fair share of blind alleys and false starts.

CNS COMPONENTS OF SOMATIC SENSATION

The central paths taken by large diameter afferents and small diameter afferents tell two stories worth attention. The first is that the path for mechanosensation and proprioception is separate and largely distinct from the path for nociception, thermoreception, and itch. These separate paths continue through the CNS to the cerebral cortex and represent one of the clearest divisions of labor seen in any sensory system. The second story is seen in the many regions of the CNS contacted by each population of axons. Here the lesson repeats a statement made at the beginning of this chapter: the somatosensory system has many tasks to perform, from the input side of motor reflexes to the higher-order functions of perception, comprehension, and emotion. To accomplish them, mechanosensory/proprioceptive and nociceptive/thermoreceptive axons send information along several pathways, performing separate functions.

An Outline of Ascending Paths to Perception

Primary afferents in the somatosensory system terminate on second-order neurons in either the spinal cord (nociceptors and thermoreceptors) or medulla (mechanoreceptors and proprioceptors) (Fig. 24.8). Second-order neurons in the spinal cord and medulla send their axons across the midline to terminate in thalamus. Convergence is kept to a minimum so that second-order mechanosensory and nociceptive neurons end in separate nuclei and subnuclei of the thalamus. In addition, axons of particular submodalities (e.g., muscle spindle afferents vs. cutaneous afferents) end in different subnuclei of the thalamus. The major thalamic nuclei involved are the various parts of the ventral posterior complex (referred to as ventral caudal in human studies). The lateral and medial subnuclei of VP, logically named VPL and VPM, are the recipients of discriminative inputs from the body and face, respectively. Both innervate the first somatosensory cortex (SI) and a third subnucleus, VPI, sends its axons to the second somatosensory area (SII).

SI and SII are recognized in the cerebral cortex of all mammals that have been examined (Fig. 24.9). Yet the organization of each appears to vary considerably across that class. If we take human cerebral cortex as a starting point, SI is traditionally divided into four structurally distinct areas named (from rostral to caudal) areas 3a, 3b, 1, and 2. The same is seen for the best-studied genus of monkeys, the macaques. What stands out in the organization of these four areas is the combination of serial and parallel processing. Parallel processing is seen in the types of thalamic neurons that innervate each area and the receptive field properties displayed by those

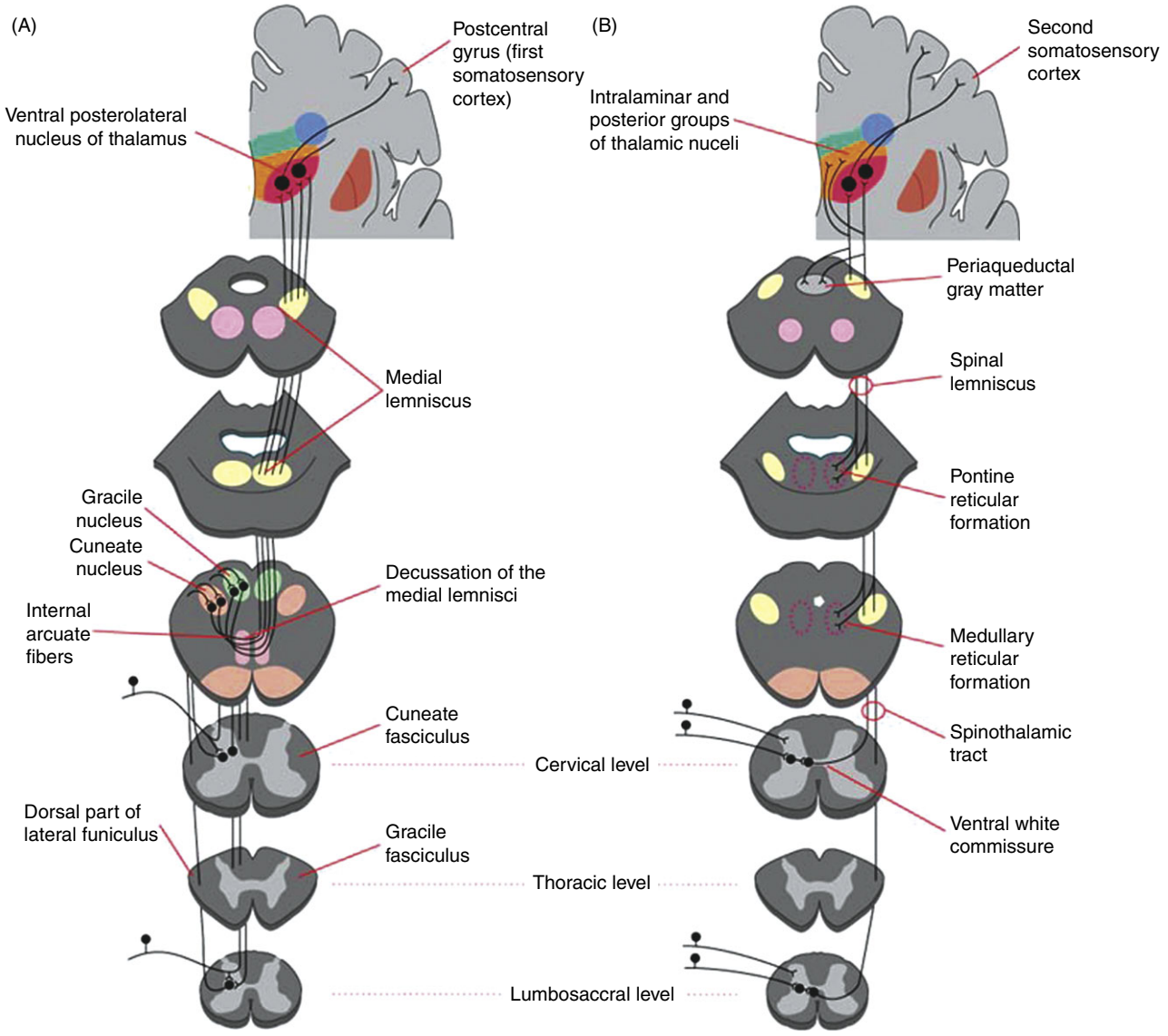


FIGURE 24.8 Anatomy of ascending somatosensory paths. (A) Organization of the dorsal column-medial lemniscal system from entry of large diameter afferents into the spinal cord to the termination of thalamocortical axons in the first somatosensory area of the cerebral cortex. An obligatory synapse occurs in the gracile and cuneate nuclei, from which second-order axons cross the midline and ascend to the ventral posterolateral nucleus of the thalamus (VPL) by way of the medial lemniscus. (B) Organization of the spinothalamic tract and the remainder of the anterolateral system. Primary axons terminate the spinal cord itself. Second-order axons cross the midline and ascend through the spinal cord and brainstem to terminate in VPL and other nuclei of the thalamus. Collaterals of these axons terminate in the reticular formation of the pons and medulla.

cortical neurons. Areas 3b and 1 are most often characterized by the response of neurons to cutaneous inputs, SAs and RAs, whereas neurons in area 3a are found to respond exclusively to stimulation of deep receptors and neurons in area 2 respond to both.

Serial processing of information is evident in both the physiology and connectivity of these areas. Area 3b is much more richly innervated by VPL and VPM than any of the other three areas of SI and, in terms of its structure and connectivity, appears much more of a primary sensory area. Comparative studies suggest that area 3b of rhesus monkeys and humans is in most ways equivalent to all of SI in other mammals and so this area is often referred to as SI Proper. The principal outputs of area 3b are directed caudally to areas 1 and 2; and in a similar fashion, area 1 sends its own rich set of axons to area 2. From these considerations, a hierarchy emerges, with area 2 at something of a pinnacle, in that it receives a direct deep receptor input from the thalamus, indirect deep input from 3a and an indirect cutaneous input from areas 3b and 1.

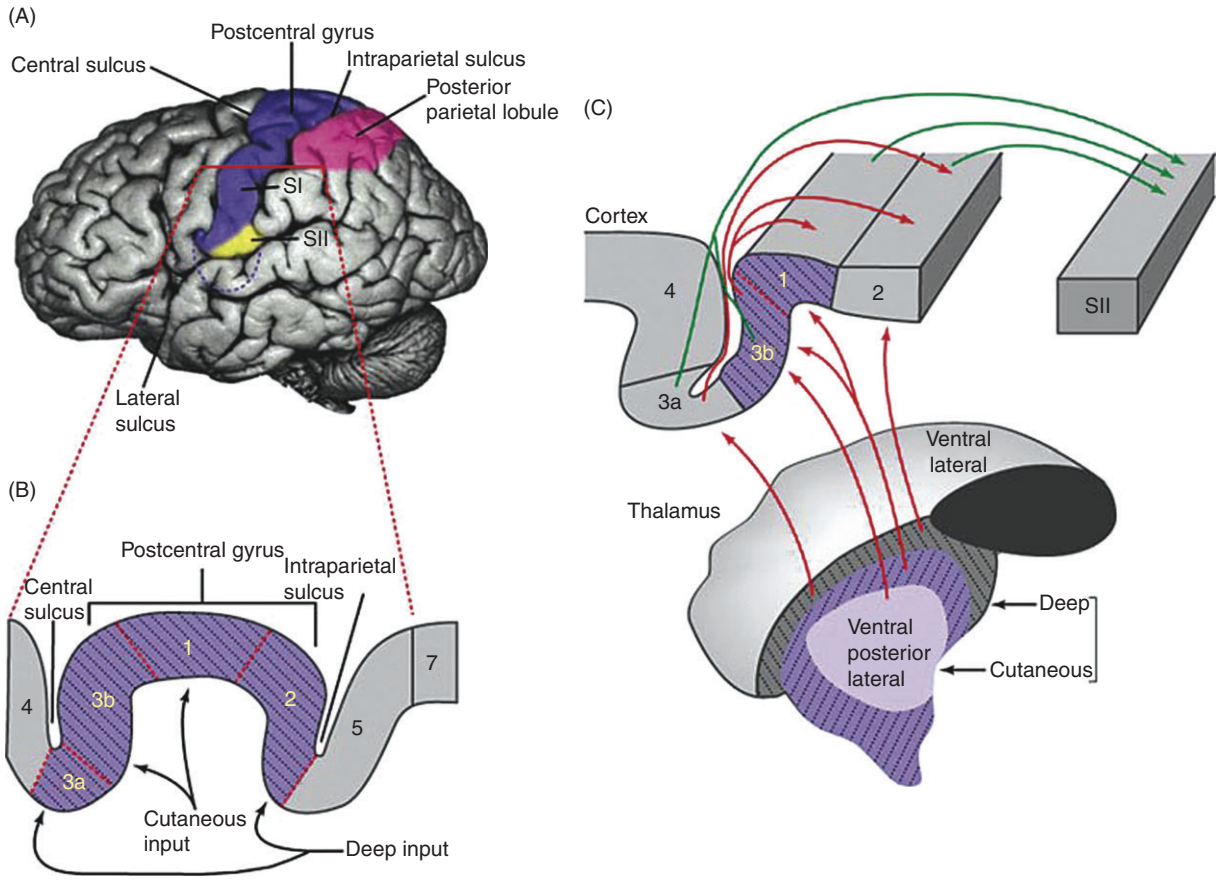


FIGURE 24.9 The first somatosensory cortex of humans and other primates is an amalgam of four areas, each of which has a complete map of the body surface. (A) SI and SII are located posterior to the central sulcus. (B) SI is shown at higher magnification. Areas are numbered according to the original scheme by Brodmann as areas 3a, 3b, 1, and 2 (from rostral to caudal in the postcentral gyrus). (C) Thalamic input from the core of VPL and VPM carries cutaneous mechanosensory information to areas 3b and 1, while a parallel route from the shell of VPL and VPM carries deep (e.g., muscle spindle) information to areas 3a and 2. Intracortical connections from area 1 to area 2 make up a route by which cutaneous information also reaches area 2. As a site of convergent deep input from thalamus and cutaneous input from area 1, area 2 is the most likely location in which a complete proprioceptive map emerges.

The output of SI as a whole is in two directions (Fig. 24.10):

1. A ventral path to SII and from there to the caudal insula, to areas of the temporal lobe, and to premotor and prefrontal cortical areas. Eventual convergence of somatosensory, auditory, and visual information in the medial temporal lobe is viewed as the path toward shape and form processing. This ventral path is vital for inserting new information into declarative memory and accessing established memories for comparison with ongoing events.
2. A dorsal path to superior parietal lobule, providing areas in that lobule with somesthetic information for control of voluntary movements, selective attention, and information about how to perform different tasks, sometimes known as the "how" pathway.

With these essential elements in mind, we will consider separately the paths for mechanosensation and nociception from the entry of primary afferents into the spinal cord to the areas of cortex in which the elements of sensation are brought together into a perceptual whole.

The Path for Mechanosensation for the Body

Dorsal columns are the principal routes for spinal mechanoreceptor axons. All axons of sensory ganglion cells enter the CNS by way of dorsal roots and the trigeminal nerves. At a gross level the spinal cord is segmented by the existence of separate dorsal root ganglia. Because the peripheral tissue innervated by any one dorsal root ganglion is restricted, the entire innervation of the body surface can be seen as a series of overlapping bands. These are the dermatomes. As a practical matter, then, a mechanical stimulus applied to a restricted region of the body surface leads to action potentials conducted in one or at most two dorsal roots.

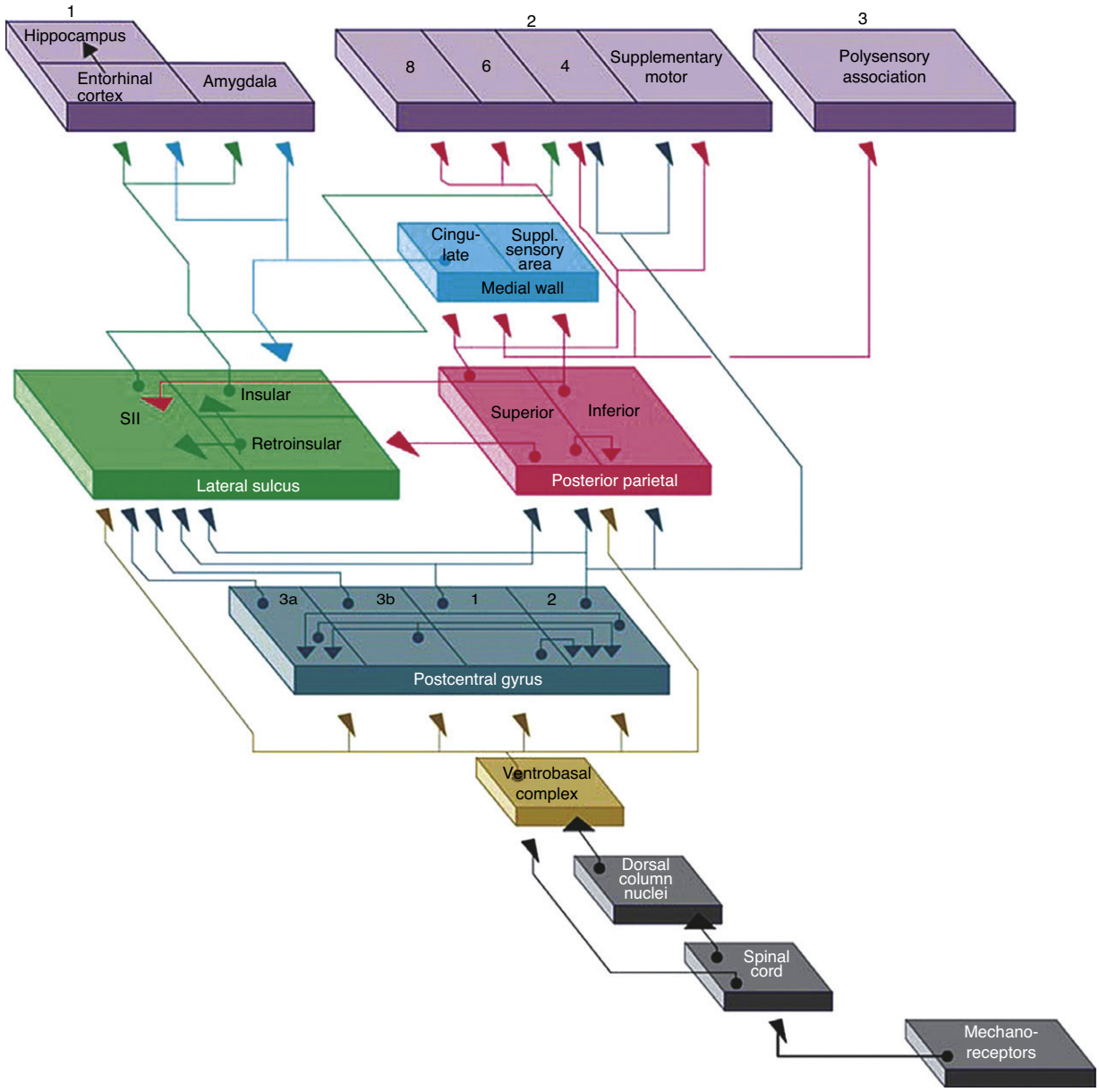


FIGURE 24.10 Schematic representation of the path taken by mechanoreceptor input to eventually reach three cortical targets. All relevant information reaches the ventrobasal complex and most is relayed to the areas of SI. From there, by steps through SII and the posterior parietal areas, somatosensory information reaches (1) the limbic system (entorhinal cortex and hippocampus), as a means for becoming part of or gaining access to stored memories; (2) the motor system (primary and supplementary motor cortex), where the continuous sensory feedback onto the motor system occurs; and (3) the polysensory cortex in the superior temporal gyrus, in which creation of a complete and abstract sensory map of the external world is thought to occur.

Mechanosensory and Nociceptive Axons Enter the Cord along Different Paths

The dorsal roots divide after they have penetrated the dura mater but before they are have entered the spinal cord itself. A medial division of large diameter, heavily myelinated axons enters at the dorsal funiculus of the cord, and the majority of those axons turn at right angles after they have entered the cord and ascend toward the brain. From level T7 of the cord to the terminal coccygeal segment, the large diameter axons enter one fiber tract called the gracile fasciculus. And as they do so they stack up as thin sheets with the earliest entering axons (coccygeal) occupying the most medial part of the funiculus and the last entering (7th thoracic) occupying the most lateral part. That same medial-to-lateral pattern continues with large axons that enter from T6 up to the first cervical segment but at those levels the axons form a separate fasciculus, called the cuneate fasciculus. Together the gracile

and cuneate fasciculi are called the dorsal columns. The orderly pattern established by the axons of the dorsal columns as they ascend means there is a body map or somatotopy to the ascending mechanosensory system. Large diameter axons ascend on the same side of the spinal cord as the one they entered to reach the lowest levels of the medulla.

Dorsal Column Nuclei

In a path toward sensory perception, the majority of axons in the dorsal columns ascend to reach the dorsal column nuclei at the junction of the spinal cord and medulla. The dorsal column nuclei are neither simple nor homogeneous. At the grossest level, each nucleus is divided into at least three regions in which axons of different mechanoreceptors terminate separately. Neurons receiving muscle spindle inputs are sufficiently distinct in some species to be given their own name: nucleus Z. All of this serves to accent the role played by subcortical somatosensory nuclei in keeping segregated the input from different functional classes of receptors.

Despite an anatomical convergence of primary afferents onto single neurons of the dorsal column nuclei, convergence and comparison of receptor input appears minimal. The coupling between a single neuron in the gracile or cuneate nucleus and a single afferent axon is extremely tight (Rowe, 2002). As a result of this coupling, a single action potential in the afferent axon reliably produces an action potential in the postsynaptic cell. That phenomenon is seen for RA, PC, and SAI afferents of the glabrous skin and hair follicle afferents of the hairy skin. These findings indicate that for a neuron in the gracile and cuneate nuclei, receptive field locations, sizes and properties are a matter of which primary afferent innervates that neuron.

The Lateral Cervical Nucleus

A cluster of mechanosensitive relay neurons starts in the dorsal horn at cervical segments 1 and 2 and extends into the caudal one-third of the medulla. This is called the lateral cervical nucleus. Innervated by mechanosensory neurons in the dorsal horn throughout all segments of the spinal cord, the neurons in the lateral cervical nucleus display predominantly cutaneous receptive fields that cover large patches of hairy skin and usually some part of glabrous skin. Most axons of cells in the nucleus decussate immediately and reach the contralateral VPL by way of the medial lemniscus. In the VPL, the cervicothalamic axons responding to low-threshold mechanical stimulation of skin or hair appear to converge with axons that arise in dorsal column nuclei. Unlike other nuclei of the somatosensory system, the lateral cervical nucleus is an inconstant feature of the mammalian nervous system. Whereas it is a robust part of the spinal cord and medulla in rodents and carnivores and has been found consistently in nonhuman primates, it is present in only half the humans studied and is well defined in very few of them. Comparison across species also shows a marked difference in the density of cervicothalamic terminations in the VPL, with the density in cats far exceeding that in rhesus monkeys. These findings lead to the assumption that functions parceled out to the lateral cervical nucleus in rodents and carnivores are sequestered in the dorsal column-medial lemniscal system of primates, and particularly of humans.

THALAMIC MECHANISMS OF SOMATIC SENSATION

Segregation of Place and Modality Continues in Thalamus

Medial lemniscal axons terminate on large neurons of the VPL and VPM. As in the dorsal column nuclei, the synaptic means to generate lateral inhibitory influences exists in the thalamus, but evidence for robust lateral inhibition is seldom found in these nuclei. Where inhibition has been detected, its strength across the neuron's receptive field matches that of the excitation so that the region of greatest inhibition is also the region of greatest excitation. Missing from this stage of somatosensory processing, then, is the synaptic means to produce contrast, by which activity produced by stimulation of one spot on the skin surface is compared with activity produced by stimulation of surrounding regions.

Neurons in the VPL and VPM display a cluster of features labeled as lemniscal properties. Static lemniscal properties are place and modality specificity. Thalamic circuits maintain these features much as they exist in the lemniscal afferents themselves. As a result, mixing of inputs is avoided in the thalamus and the specific properties of where on the body surface something has occurred and what exactly has occurred remain anatomically distinct. Dynamic properties characteristic of these cells include the ability to follow a rapid train of stimuli. Here the synaptic organization of the thalamus favors the ability of a thalamic neuron to behave much like the axon that innervates it.

Anatomical and physiological studies demonstrate that as lemniscal afferents terminate in the VPL, they do so as clusters of axon terminals elongated in the rostro-caudal dimension. They are referred to as rods. Rods appear to be the major organizing principle to somatosensory thalamus, as lemniscal axons carrying information of the same type (e.g., SAI input) from the same part of the body terminate in a rod-like formation. On the output side, thalamocortical neurons that send axons to a 1-mm-wide patch in SI occupy the same sort of formation. Thus, thalamic rods are the anatomical basis for the most salient

feature of somatosensory thalamus: the strict segregation of place- and modality-specific responses.

THE PATH FROM NOCICEPTION TO PAIN

Spinal Cord Pathways

In stark contrast to the pattern of input by mechanosensory axons, nociceptive axons terminate in the dorsal horn of the spinal cord. These lightly myelinated and unmyelinated axons enter the cord by way of a lateral division of the dorsal root and divide to send branches up and down the cord for a segment or two. The tract they form is a cap on the surface of the dorsal horn called Lissauer's tract. Nociceptive axons terminate, therefore, on dorsal horn neurons across four or five segments, and a single dorsal horn neuron is innervated by nociceptors that cover a broad swath of the body or limb surface. Second-order neurons innervated by the nociceptive axons give rise to axons that decussate and ascend in the anterolateral quadrant of white matter to reach the brainstem and thalamus.

Spinal cord circuits for nociception form two ascending routes with distinct functions (Fig. 24.11). Recall that the peripheral basis for first and second pain is the division of nociceptors into relatively fast-conducting Aδ axons and slow-conducting C fibers (Braz et al., 2005).

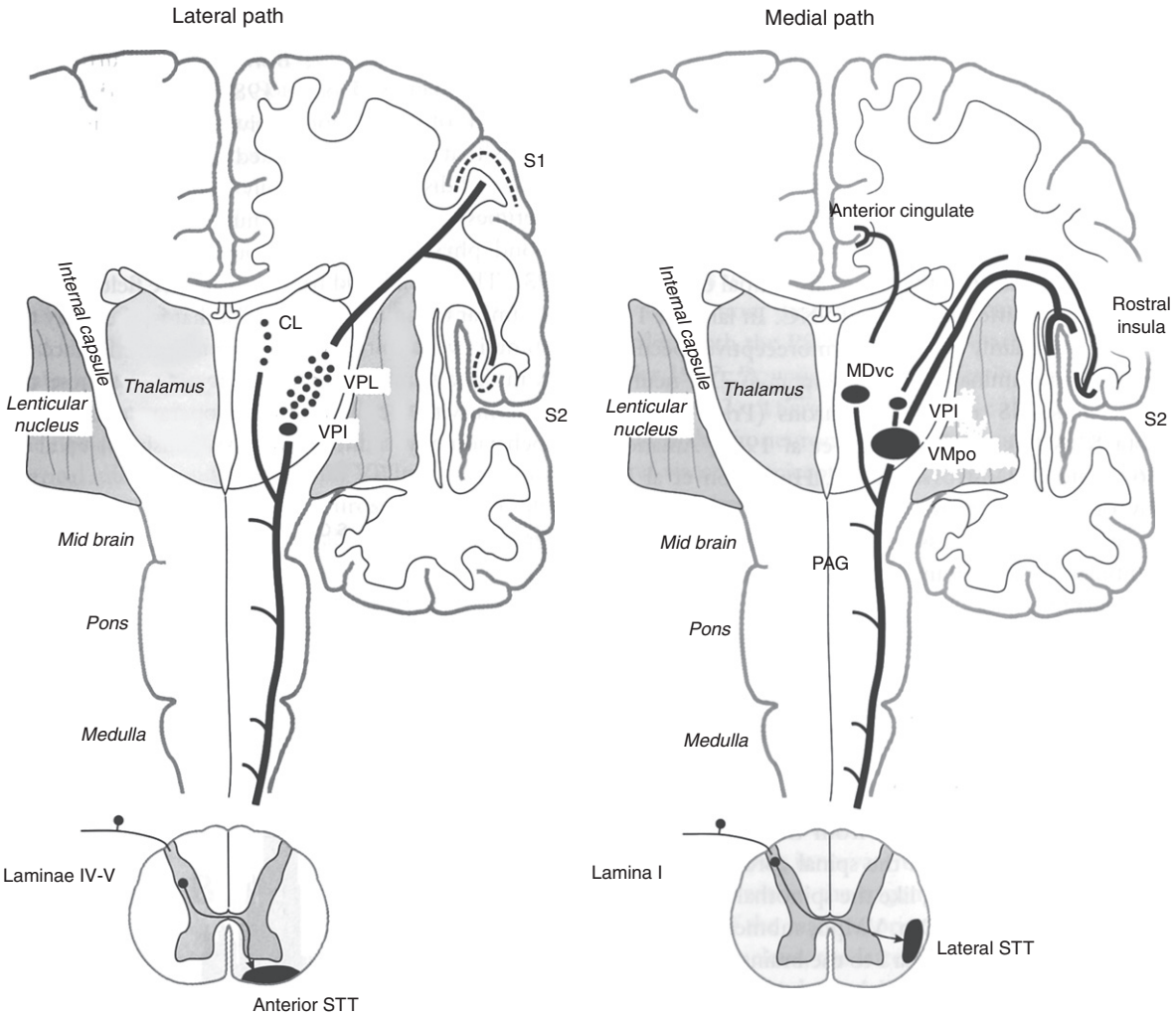


FIGURE 24.11 The medial path taken by spinothalamic neurons in lamina I (driven by C fibers) differs from the lateral path taken by neurons in laminae IV and V (driven by C and Aδ nociceptors and Aβ mechanoreceptors). Anterior spinothalamic tract axons are given off by the deeper neurons and terminate in lateral thalamus (VPL and VPI and the centrolateral nucleus). Lateral spinothalamic tract axons given off by lamina I neurons innervate medial thalamus, including the ventral caudal division of the mediodorsal nucleus (MDvc) and a region in posterolateral thalamus. In this figure, the region is labeled a separate, nociceptive/thermoreceptive-specific nucleus called VMpo, but several lines of evidence indicate lamina I neurons also innervate VPL and VPM, as well as VPI. From this varied thalamic innervation, nociceptive information reaches SI and SII for discriminative aspects of pain and temperature and the anterior cingulate and rostral insula for the affective, punishing aspects of pain.

Yet the discriminative versus agonizing components of pain is a CNS construct, produced by the difference in central circuits driven by Aδ and C fibers. To appreciate that difference, it is best to start with the population of spinothalamic neurons.

Spinothalamic Neurons

Spinal cord neurons that directly innervate thalamus (therefore, spinothalamic neurons) occupy several laminae in rhesus monkey spinal cord. The major populations, however, are found in the most superficial layer of dorsal horn (lamina I) and a second region deeper in the dorsal horn (lamina V). Nociceptive innervation of these two populations differs fundamentally. Lamina I neurons are innervated predominantly by C fibers, both directly and indirectly by way of excitatory interneurons in the immediately adjacent superficial half of lamina II. Theyalso receive an indirect innervation from Aδ fibers by those same interneurons and can be thought of as nociceptive-specific cells. Lamina V neurons, on the other hand, are innervated predominantly by Aδ fibers that end deep in lamina II on a population of excitatory interneurons. Many of these deeper spinothalamic neurons also receive a convergent input from mechanosensory axons (Aβ axons). Because they respond to both nociceptive and mechanosensory stimulation, these neurons are often referred to as wide dynamic range cells.

Ascending Paths to the Thalamus

Axons from both lamina I neurons and laminae IV/V neurons decussate and enter fiber tracts in the anterolateral quadrant of spinal cord, but the tracts they enter differ from one another in location and termination (Fig. 24.11). Neurons of laminae IV and V enter the anterior spinothalamic tract and terminate in lateral parts of the thalamus, including the VPL and VPI, and an intralaminar nucleus, the central lateral nucleus. The innervation of ventral posterior thalamus does not mean both nociceptive and mechanosensory information converge in the thalamus. The two remain segregated as medial lemniscal axons terminate on groups of large neurons, whereas spinothalamic axons innervate clusters of smaller neurons. Response properties of the smaller neurons in the VPL are much like those of the wide dynamic range neurons that innervate them, both in the type of stimuli to which they respond and the size of their receptive fields.

Nociceptive specific neurons of lamina I enter the lateral spinothalamic tract and terminate in several nuclei of the thalamus, including two that receive few deep lamina inputs (Fig. 24.11). Targets in which convergence of lamina I and lamina V inputs are strongly suspected or known to occur include the nucleus of the ventral posterior complex. Although the most compelling data in support of convergence has been reported for VPI, several careful studies have documented termination of lamina I inputs in the clusters of small neurons in VPL and VPM. This represents what has become a traditional view of the nociceptive pathway, that discriminative pain is driven by nociceptive inputs to both lamina I and lamina V, through a relay in the ventral posterior nuclei, and that nociceptive information reaches SI and SII (Perl and Kruger, 1996).

Outside the ventral posterior complex are two sites of spinothalamic terminations predominantly from lamina I (Fig. 24.11). One of these is a subnucleus of the mediodorsal nucleus (Mdvc) in which spinothalamic terminations are surprisingly rich. The other is posterior to the classically drawn borders of VP and includes the medial nucleus of the posterior group (POm). Neurons in this region are innervated by spinothalamic axons and display the same chemical signature (immunoreactivity for the calcium-binding protein, calbindin) as spinothalamic-recipient neurons of VPL and VPM (Craig and Dostrovsky, 1999).

Considerable disagreement exists about the spinothalamic innervation in posterolateral thalamus of monkeys and humans. One part of the disagreement deals with the proportion of lamina I afferents that end inside the bounds of the VPL and VPM and those that end caudal and medial to it. A conservative reading of the literature leads us to recognize spinothalamic innervation of four regions, each with its own cortical target. Spinothalamic terminations in the VPL and VPM are relayed to SI and those in VPI to SII. Together, they can be seen as a lateral path to first or discriminative pain. A second medial path to anterior cingulate by way of MDvc and to the insula by neurons of VMpo appears to be the central route for second or punishing pain.

SI and Pain

Nociceptive responses in SI have been difficult to record, and where neurons responding to noxious stimuli have been encountered, their location (e.g., area 3a) has not been easy to fit into a general scheme of what these areas do. Nevertheless, ablation studies in monkeys show SI's involvement in pain perception, and studies of human cerebral cortex regularly show that SI responds to stimuli judged to be painful. Perhaps the most compelling story can be told for area 1 of SI in monkeys, where evidence exists of clustered organization for nociceptive neurons (Kenshalo et al., 2000). Yet these neurons do not form columns. They are largely confined to layer IV of that area and are intermixed with mechanosensory neurons. These data suggest that nociceptive and mechanosensory inputs converge in area 1, perhaps as a means to accurately locate the source of pain.

The Human Axis of Pain

Functional imaging studies of the human brain indicate four areas of cerebral cortex are active during the application of a painful stimulus. Activation of SI and of SII occurs

as part of a discriminative component to painful stimuli. A subject's ability to report the location and to grade the intensity of a stimulus is correlated with activity in these areas. Two other areas appear tied to the cognitive and emotional content of pain. These are the rostral half of the anterior cingulate gyrus and the rostral insula, both of which display elevated activity during the anticipation of painful stimuli and the infliction of pain on a loved one (empathetic pain). And it is these areas that show a reduction in activity when the administration of a placebo produces reports of lessened pain. As a group, then, studies of the cortical representation of pain point to a distributed network of four main areas (SI, SII, rostral anterior cingulate, and rostral insula) in which the noxious stimulus applied to some peripheral region becomes a painful sensation with both discriminative and punishing components.

Nonperceptual Elements of Nociception

Several paths are taken by first- and second-order nociceptor axons that reach neither spinothalamic neurons nor regions of the thalamus. Targets of these axons include spinal interneurons that mediate the withdrawal and crossed-extensor reflexes; the pontine reticular formation, where the startle reflex is generated; and the midbrain periaqueductral gray (PAG) as a means for controlling pain. The last of these is a well-studied mechanism. Neurons of the PAG are innervated by second-order nociceptive neurons in laminae V. They provide PAG neurons with an indication of the source and intensity of nociceptive input. By way of a relay from serotonin neurons in nucleus raphe magnus and norepinephrine neurons in the locus coeruleus, activity in PAG drives a population of interneurons in the lamina II of the spinal cord. Two characteristics of those interneurons are worthy of consideration. First, they end not only on the populations of spinothalamic neurons but also on the axon terminals of C and Aδ axons. Through postsynaptic inhibition these inhibitory interneurons are able to suppress the response of spinothalamic cells and through presynaptic inhibition they are able to suppress release of the neurotransmitter, glutamate, from primary nociceptive afferents. The combination of pre- and postsynaptic inhibition reduces activity in the population of spinal neurons that carries nociceptive information to the brain. Second, the inhibitory neurons release the pentapeptide, met-enkephalin as a neurotransmitter. Met-enkephalin is an opioid peptide that binds to a member of the opiate receptor family. Its actions are mimicked by the administration of morphine and synthetic opiates accounting for at least part of the analgesic properties those compounds possess.

Pain is such a rich experience with an obvious impact on the well-being of a person that unusual phenomena associated with pain have been studied in considerable depth. Among the best studied are referred pain, secondary hyperalgesia/allodynia, and phantom limb pain. Each has been explained by referring to the circuitry and chemistry of normal nociceptive processing. Referred pain is the experience in which noxious stimuli in the viscera (e.g., ischemia of cardiac muscle) is felt as pain in a peripheral location such as the shoulder. A classic explanation for referred pain notes the widely branching nature of C and Aδ fibers as they enter the cord, so that lamina I neurons innervated by nociceptors of the shoulder are also innervated by nociceptors of the pericardium. When the latter are driven to fire a series of action potentials the percept is one of a more common occurrence, shoulder pain.

Central or secondary hyperalgesia and allodynia are related phenomena that occur in response to synaptic plasticity at the level of the spinal cord. Secondary hyperalgesia is a phenomenon in which a greater degree of pain is felt upon application of a noxious stimulus to a place other than the wound site. Allodynia, by contrast, is the perception of pain when the stimulus, itself, is not noxious. Each occurs when the application of an intense or prolonged noxious stimulus leads in spinal circuits to a rearrangement in synaptic strength, much like long-term potentiation. With secondary hyperalgesia, the potentiation is homosynaptic, leading to a greater postsynaptic response in spinal neurons to a noxious stimulus. Plasticity with allodynia is thought to be heterosynaptic. Under normal circumstances, activity in Aβ fibers to a benign mechanical stimulus (e.g., the movement of a cotton swab across the skin) modulates a response to noxious stimuli but fails to drive spinothalamic cells. Yet when the synapses formed by the Aβ fibers are potentiated in response to an intense or prolonged barrage of C fiber activity, spinothalamic neurons are driven by that mechanical stimulus. The result is perception of pain where only nonnoxious mechanical stimulation has occurred.

Unlike allodynia, phantom limb pain appears to involve synaptic plasticity at each of several levels in the somatosensory system. Removal of a digit or limb in experimental animals and in humans leads to a robust functional plasticity. Surgical removal of digits produces a short-lived silent zone in the somatotopically appropriate part of SI, but quickly thereafter, the previously silent region becomes responsive to tactile stimulation. The newly acquired response is to adjacent body parts, as the activity evoked in neighboring receptors fills in the zone where the cortex had been silent. Subsequent work has shown that following amputation of a digit or limb, this process of filling in takes place through synaptic changes in the spinal cord, medulla and thalamus, as well as in cerebral cortex. It is assumed that phantom limb pain works much the same way. Thus, stimulation of the shoulder and chest in amputees drives regions of cerebral cortex that had been responsive to stimulation of an arm prior to its amputation. The result is the percept of arm pain even in the absence of an arm (Box 24.1).

BOX 24.1

THE TRIGEMINAL SYSTEM

Mechanoreceptive, nociceptive, and thermoreceptive afferents for the face have their cell bodies in the pair of trigeminal ganglia (Fig. 24.12). The central processes of trigeminal ganglion cells enter the mid-pons as the trigeminal nerve. In many ways, the central trigeminal system is organized along parallel lines with the spinal somatosensory system.

Three nuclei make up the somatosensory part of the trigeminal system. The largest of these is the principal or main sensory nucleus, the cell bodies of which are in mid-pons, at the level of entry for the trigeminal nerve. Acting much like the dorsal column nuclei, the principal sensory nucleus is innervated by large diameter afferents of the ipsilateral half of the face. Its neurons respond to skin indentation and to vibrotactile stimuli. Most of these neurons send their axons across the midline in the pons, where they join the fibers of the medial lemniscus. The trigemino-thalamic axons ascend in the most medial part of the lemniscus (next to the axons that carry information about the hand) and terminate in the VPM.

The spinal trigeminal nucleus is an elongated nucleus, split into three subnuclei distinguished geographically by reference to the obex. Pars oralis occupies the lower pons and upper medulla to the level of the obex. It gives way to the pars interpolaris at the obex and then to pars caudalis

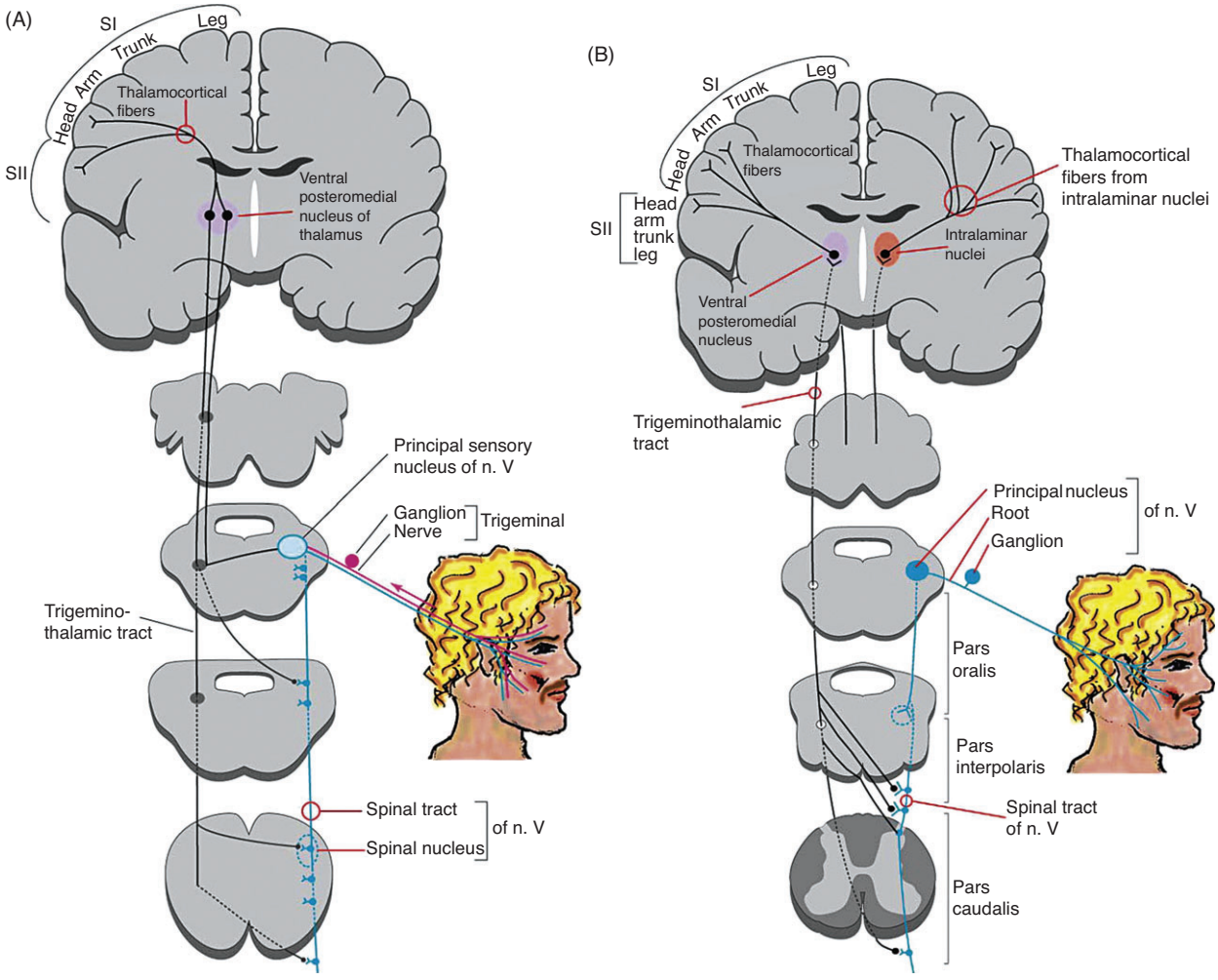


FIGURE 24.12 Sensory components of the trigeminal system. (A) Path for discriminative touch. Large diameter afferents from the face innervate second-order neurons in the spinal trigeminal nucleus (pars oralis) and the principal sensory nucleus. Neurons in these nuclei give rise to axons that cross the midline, ascend in the trigeminothalamic tract, and terminate in the ventral posteromedial (VPM) nucleus of the thalamus. (B) Path for pain and temperature in the trigeminal system. Small diameter afferent axons descend in the spinal trigeminal tract and terminate in the pars caudalis of the spinal nucleus. Second-order axons cross the midline and ascend to the thalamus.

BOX 24.1 (*cont'd*)

through the lower medulla and into the first two cervical segments of the spinal cord. Borders between these subdivisions are hardly distinct as afferents innervating a restricted part of the face may end across two or all three parts of the spinal trigeminal nucleus. The traditional subdivision of the spinal nucleus into functional units accents the mechanosensory functions of pars oralis, the deep receptor functions of pars interpolaris and the nociceptive and thermoreceptive functions of pars caudalis. In this context, the relay of input from pars caudalis to VPM is viewed as the equivalent of the spinothalamic system for the face. Yet studies of C fibers that innervate tooth pulp find these afferents terminate in a long, continuous sheet from the caudal half of the principal nucleus to the caudal-most aspect of pars caudalis. These data indicate a great deal more intermixing of mechanosensory neurons and nociceptive neurons takes place in the trigeminal system than is seen for the spinal somatosensory system.

CORTICAL REPRESENTATION OF TOUCH

Neurons of SI Are the First in the Somatosensory System to Show Clear Signs of Lateral Inhibition

Between the level of the peripheral afferent and the level of the cerebral cortex, the somatosensory system keeps information from individual afferents from coming together. Even though the synaptic means exist to generate lateral inhibition in the dorsal column nuclei and in thalamus, the response of single neurons in those regions appears very much like the response of the axon that drives it best. Neurons in SI, however, display true inhibitory surrounds. For area 3b in monkeys, such surrounds are one-third larger than the central excitatory region and can occupy one or more (but rarely all four) sides of the excitatory center. Many neurons in area 3b also exhibit a strong temporal component to their response. Spatial inhibition maintains the response selectivity of neurons when objects scanned across the skin surface at different speeds (as in reading Braille); receptive fields of neurons in area 3b are velocity insensitive. Temporal inhibition performs a parallel function by producing a progressive increase in response rate to surface elements that are scanned more quickly. The lag between the initial excitatory response and the delayed inhibitory one liberates more activity when the velocity of scanning increases. This more than makes up for the reduction in contact time between peripheral receptor and surface feature that accompanies any increase in scanning velocity.

Orientation Selectivity

As a result of circuits in cerebral cortex, the roughly circular receptive fields that enter SI by way of thalamocortical axons become orientation-selective responses in 70% of cortical neurons. The synaptic mechanism for generating a receptive field tuned to the orientation of a stimulus could be convergence of excitatory inputs, much like the mechanism proposed originally for orientation tuning in visual cortex. Yet examination of spatial-temporal response of neurons in SI indicates orientation selectivity is best explained by the presence and spatial distribution of inhibitory regions.

Serial Processing in SI

Area 3b is considered an early step in the cortical processing of tactile information. Much of the data supporting that conclusion come from studies that show ablation of area 3b produces a general decline in tactile discrimination behavior whereas ablation in other areas of SI leads to functional loss that is more selective. These deficits include difficulties in discriminating the texture of a surface following ablation of area 1 and deficits in three-dimensional form perception with ablation of area 2. From these data we conclude that areas of SI are organized as a hierarchy. A similar conclusion is seen in physiological properties of SI neurons as receptive fields become larger and their properties become more complex in going from area 3b to one of the more caudal somatosensory areas (Pons et al., 1992). One example is selectivity for the direction of object movement across the skin. Neurons in area 3b show no sign of direction selectivity, but in areas 1 and 2 this property emerges from the cortical circuitry (Luna et al., 2005).

Columnar Organization

Responses of like type are organized as columns in SI. For SI in macaques one obvious place for columnar processing is in the RA and SA responses. Neurons in layer IV of monkey area 3b, directly innervated by thalamocortical axons, show strong specificity for either RA or SA inputs (Sur et al., 1984). Yet superficial to layer IV, almost all neurons display slowly adapting responses. In other words, the response of neurons outside layer IV makes it appear as though rapidly adapting inputs are incorporated into a slowly adapting output. Nevertheless neurons in areas 3b and 1 are clearly involved in the discrimination of vibrotactile stimuli at low frequencies, such as those

carried into the CNS by RA afferents (Meissner's corpuscles). Response of SI neurons in a macaque is tied to the frequency of a stimulus applied to the monkey's fingertips. And so his ability to discriminate between two stimuli of different frequencies is only as good as an SI's neuron's ability to signal that difference.

The Role of SII in Somatic Sensation

Ablation studies reveal three features of SII's role in somesthesis:

1. Removal of all SII leaves a macaque incapable of discriminating the shape and texture of a tactile stimuli. These data show any or all subdivisions of SII play a pivotal role in tactile behavior.
2. Removal of SII has no effect on the response of neurons in SI. So, even though SII provides feedback projections to SI, the dominant input to SI comes from neurons of VPL and VPM.
3. Removal or cooling of SI eliminates the tactile responses in SII.

These results extend to modality-specific loss in SII following ablations restricted to one area in SI (removal of area 1 disrupts response to stimulus texture in SII). These observations place SII at a higher level on a series of steps that leads from SI into the medial temporal lobe, where access to memory is achieved

Several Body Representations in SII

At first glance, the division of labor that is a hallmark of SI—its division into four areas—contrasts strongly with the presence of a single area called SII. Recent functional and anatomical studies show, however, that the region in the upper bank of the lateral fissure that normally defines SII in macaques is divided into multiple functionally distinct areas (Fitzgerald et al., 2006). In monkeys three separate regions have been described, referred to as SIIa, SIIc, and SIIp (some researchers prefer the terms SII, PV, and PR; Krubitzer et al., 1995). Anatomical studies in humans suggest there are also four separate regions called OP1–OP4, yet how the human studies relate to the monkey studies is not known. So a modern view of SII indicates it is a complex of at least three functionally distinct areas. The anterior and posterior portions (SIIa and SIIp) are involved in integrating proprioceptive inputs with cutaneous input as a means for representing the size and shape of objects held in the hand. In addition, a central field (SIIc) responds to cutaneous inputs and is important for processing information about 2D shape and texture discrimination.

SII Receptive Fields

Receptive field properties of neurons in macaque SII support the conclusions that SII is part of a serial processing scheme. Neurons in SII are selective for more complex features of a stimulus than are neurons in SI. Moreover, orientation selectivity is seen in fewer than one out of three neurons in SII as receptive fields of single neurons grow to encompass two or more fingers. Neurons such as these appear to represent a sparse code for objects such as the edge of a keyboard or large curved shapes that span several digits.

SII and Attention

Psychophysical studies show that attention is a spotlight that can move freely across somatosensory functions: it has finite boundaries and improves performance on many tactile discrimination tasks. Single unit studies show SII to be one of the major targets of attention as the response of individual neurons changes with shifts in attention. Mechanisms by which attention appears to exert its effects include a change in the response rate—90% of SII neurons in macaques increase their response to a tactile stimulus when the animal attends to the stimulus—and a shift across a population of neurons so that they become active at the same time. Pairs of neurons in macaque SII become more synchronous in their response to a tactile stimulus when an animal pays attention to that stimulus. As the task becomes more difficult, when the difference between two stimuli is deliberately reduced, the degree of synchrony increases by a factor of four. These data show that not only is SII a nodal point for the discrimination of form and texture, but it is also a site at which cognitive mechanisms like selective attention are likely to exert their greatest influence.

At a more global level, SII can be viewed as a gateway through which somatosensory information reaches the medial temporal lobe. There, ongoing events along the skin surface are deposited into long-term memory and compared with previous events stored in memory. This temporal path in somatic sensation underlies object identification—the kind of processing that permits us to pull out from our pockets the right set of keys as we approach our front doors in the middle of the night. The next logical act, inserting the right key into the right lock, demands sensory guidance of movement, and that is the role of the output from SI to posterior parietal lobe. Proprioceptive inputs guide our arms and hands to the correct configuration (the height of the arm, the angle of the hand and the extension of the fingers), while central processing of cutaneous mechanoreceptors, particularly Pacinians, informs us whether the key has struck the side of the lock or moved smoothly into it. That sort of sequence of object recognition and guidance of movement repeats itself dozens, perhaps hundreds, of times a day for all of us, in each case operating efficiently and seamlessly thanks to the central processing of a complex sensory input.

References

Braz, J. M., Nassar, M. A., Wood, J. N., & Basbaum, A. I. (2005). Parallel "pain" pathways arise from subpopulations of primary afferent nociceptor. *Neuron, 47,* 787–793.

Craig, A. D., & Dostrovsky, J. O. (1999). Medulla to thalamus. In P. D. Wall & R. Melzack (Eds.), *Textbook of Pain* (4th ed., pp. 183–214). Philadelphia: Churchill Livingstone.

Fitzgerald, P. J., Lane, J. W., Thakur, P. H., & Hsiao, S. S. (2006). Receptive field (RF) properties of the macaque second somatosensory cortex: RF shape, size and somatotopic organization. *Journal of Neuroscience, 26,* 6485–6495.

Friedman, R. M., Khalsa, P. S., Greenquist, K. W., & LaMotte, R. H. (2002). Neural coding of the location and direction of moving object by a spatially distributed population of mechanoreceptors. *Journal of Neuroscience, 22,* 9556–9666.

Iggo, A., & Muir, A. R. (1969). The structure and function of a slowly adapting touch corpuscle in hairy skin. *Journal of Physiology London, 200,* 763–796.

Johnson, K. O. (2001). Neural basis for haptic perception. In H. Pashler & S. Yantis (Eds.), *Stevens handbook of experimental psychology* (3rd ed., Vol. 1: Sensation and Perception, pp. 537–583). New York, NY: Wiley.

Johnson, K. O., Hsiao, S. S., & Yoshioka, T. (2002). Neural coding and the basic law of psychophysics. *The Neuroscientist, 6,* 111–121.

Kenshalo, D. R., Iwata, K., Sholas, M., & Thomas, D. A. (2000). Response properties and organization of nociceptive neurons in area 1 of monkey primary somatosensory cortex. *Journal of Neurophysiology, 84,* 719–729.

Krubitzer, L. A., Clarey, J., Tweedale, R., Elston, G., & Calford, M. B. (1995). A redefinition of somatosensory areas in the lateral sulcus of macaque monkeys. *Journal of Neuroscience, 15,* 3821–3839.

Lewin, G. R., & Moshourab, R. (2004). Mechanosensation and pain. *Journal of Neurobiology, 61,* 30–44.

Luna, R., Hernandez, A., Brody, C. D., & Romo, R. (2005). Neural codes for perceptual discrimination in primary somatosensory cortex. *Nature Neuroscience, 8,* 1210–1219.

Perl, E. R., & Kruger, L. (1996). Nociception and pain: Evolution of concepts and observations. In L. Kruger (Ed.), *Touch and pain* (pp. 180–212). New York, NY: Academic Press.

Pare, M., Behets, C., & Cornu, O. (2003). Paucity of presumptive Ruffini corpuscles in the index finger pad of humans. *Journal of Comparative Neurology, 456,* 260–266.

Phillips, J. R., Johansson, R. S., & Johnson, K. O. (1990). Representation of Braille characters in human nerve fibers. *Experimental Brain Research, 81,* 589–592.

Pons, T. P., Garraghty, P. E., & Mishkin, M. (1992). Serial and parallel processing of tactual information in somatosensory cortex of rhesus monkeys. *Journal of Neurophysiology, 68,* 518–527.

Rowe, M. J. (2002). Synaptic transmission between single tactile and kinesthetic sensory nerve fibers and their central target neurons. *Behavioural Brain Research, 135,* 192–212.

Sandküler, J. (2009). Models and mechanisms of hyperalgesia and pain. *Physiological Reviews, 89,* 707–758.

Sur, M., Wall, D. T., & Kaas, J. H. (1984). Modular distribution of neurons with slowly adapting and rapidly adapting responses in area 3b of somatosensory cortex in monkeys. *Journal of Neurophysiological, 51,* 724–744.

Tominaga, M., Caterina, M. J., Malmberg, A. B., Rosen, T. A., Gilbert, H., Skinner, K., Julius, D., et al. (1998). The cloned capsaicin receptor integrates multiple pain-producing stimuli. *Neuron, 21,* 531–543.

Suggested Readings

Apkarian, A. V., Bushnell, M. C., Treede, R. -D., & Zubieta, J. -K. (2005). Human brain mechanisms of pain perception and regulation in health and disease. *European Journal of Pain, 9,* 463–484.

Clark, F. J., & Horch, K. W. (1986). Kinesthesia. In K. R. Boff, L. Kaufman & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 1: Sensory processes and perception, pp. 1–62). New York, NY: Wiley and Sons.

Hsaio, S. S., & Gomez-Ramirez, M. (2012). Neural mechanisms of tactile perception. In I. B. Weiner (Ed.), *Behavioral Neuroscience* (pp. 348–412). Hoboken, NJ: John Wiley & Sons.

Kaas, J. H. (2000). Organizing principles of sensory representations. *Novartis Foundation Symposium, 228,* 188–198.

Mountcastle, V. B. (2005). *The sensory hand*. Cambridge, MA: Harvard University Press.

Willis, W. D., & Westlund, K. N. (1997). Neuroanatomy of the pain system and of the pathways that modulate pain. *Journal of Clinical Neurophysiology, 14,* 2–31.

Stewart Hendry and Steven Hsiao

This page intentionally left blank

CHAPTER

25

Audition

The auditory system detects sound and uses acoustic cues to identify and locate sounds in the environment. The auditory system shares functional and evolutionary similarities with other mechanoreceptive systems, such as the vestibular system and the lateral line system of lower vertebrates. All of these systems use the same type of receptor cell—the hair cell—and all are specialized to detect an external stimulus that eventually causes the stereocilia of the hair cells to be displaced. Unlike other mechanoreceptive systems, the auditory system is sensitive to sound. This chapter explores the characteristics of the auditory system that make it sensitive to sound and that make it capable of localizing a sound source. The generalized mammalian auditory system is emphasized, but examples are also taken from studies of animals with specialized auditory systems, such as those of bats and owls, which have greatly advanced our knowledge of audition.

EXTERNAL AND MIDDLE EAR

The peripheral auditory system is divided into the external, middle, and inner ears (Fig. 25.1). The external ear is composed of the pinna and external auditory canal. These structures convey sound to the middle ear, but not without influencing it. This influence emphasizes and deemphasizes certain sound frequencies, resulting in peaks and notches in the transmitted sound spectrum. Because positions of the peaks and notches depend on the location of the sound source, they provide information about location, even when using only one ear (monaural sound localization). Such information is especially important for determining the elevation of a sound source because spectral peaks and notches are one of the few cues that depend on elevation.

Mechanosensitive organs in lower vertebrates, such as the lateral line of fish, are sensitive to vibrations in aquatic environments. When vertebrates evolved onto land, they faced the problem of converting sound in air to sound in fluid, as the inner ear of vertebrates remains fluid filled. Sound at an air–fluid interface is almost completely reflected back into the air rather than being transmitted into the fluid. The function of the middle ear is to ensure efficient transmission of sound from air into the fluid of the inner ear (reviewed by Geisler, 1998). The middle ear begins at the tympanic membrane (eardrum), continues with the three middle ear ossicles (malleus, incus, and stapes), and ends at the footplate of the stapes, which contacts the inner ear fluids at the oval window of the cochlea (Fig. 25.1). There are several ways in which the middle ear increases the transmission of sound into the inner ear. Most importantly, because the area of the eardrum is larger (by about 35 times) than the area of the stapes footplate, there is a corresponding increase in pressure from the eardrum to the stapes footplate. Additionally, there may be small contributions from a lever action of the ossicles and a buckling motion of the tympanic membrane, both increasing the force applied to the footplate. Together, these mechanisms provide a pressure gain of about 25 to 30 dB in the middle frequencies over what would be achieved by sound striking the oval window directly. This amount of gain means that for middle frequencies, most of the acoustic energy that strikes the eardrum is transmitted through the middle ear. For lower and higher frequencies, however, energy is lost. When sound conduction through the middle ear is compromised, a patient has a conductive hearing loss. One disease that causes a conductive hearing loss is otosclerosis, in which bony growths around the stapes cause it to adhere to surrounding bone and lessen its transmission

© 2013 Elsevier Inc. All rights reserved.

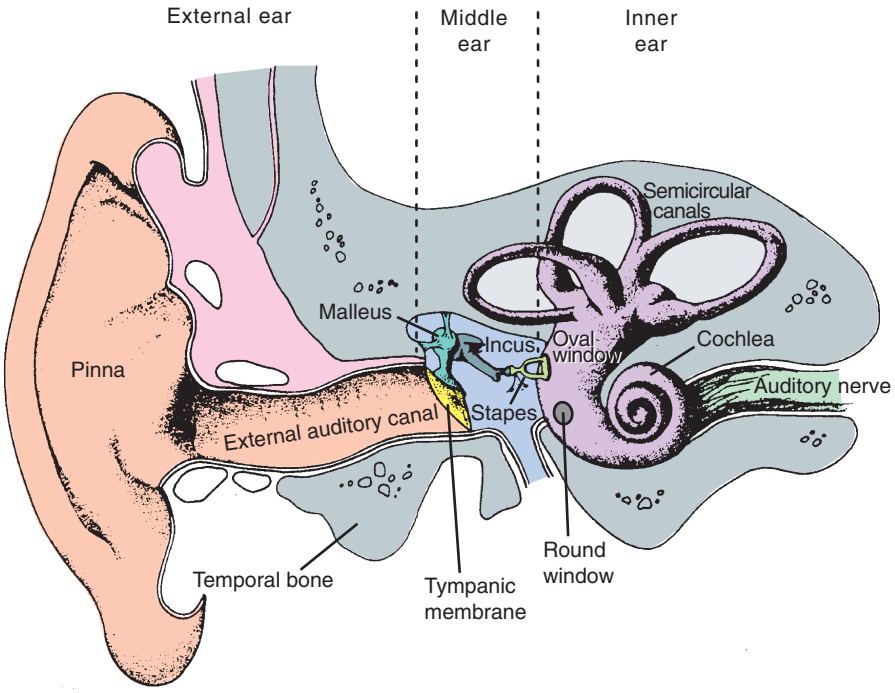


FIGURE 25.1 Drawing of the auditory periphery within the human head. The external ear (pinna and external auditory canal) and the middle ear (tympanic membrane or eardrum, and the three middle ear ossicles: malleus, incus, and stapes) are indicated. Also shown is the inner ear, which includes the cochlea of the auditory system and the semicircular canals of the vestibular system. There are two cochlear windows: oval and round. The oval window is the window through which the stapes conveys sound vibrations to the inner ear fluids. *From Lindsay and Norman (1972).*

of vibration. A surgical procedure called a stapedectomy usually restores hearing by replacing the native stapes with a prosthetic one.

THE COCHLEA

The inner ear is located deep within the head (Fig. 25.1). The inner ear contains the cochlea, which is the sensory end-organ for the auditory system. It also contains the utricle, saccule, and cristae of the three semicircular canals, which are the sensory end-organs for the vestibular system. The word *cochlea* comes from the Greek word *kokhlias*, meaning "snail," as the cochlea is coiled like the shell of a snail. In most species, the long spiraled tube of the cochlea has two to four turns, which are visible in cross section (Fig. 25.2A). The sensory organ, the organ of Corti, contains the receptor cells (hair cells) and supporting cells. The organ of Corti rests on the basilar membrane and is covered by the tectorial membrane (Fig. 25.2B). Hair cells are of two types: inner hair cells and outer hair cells. Their names are derived from the position of the hair cells along the cochlear spiral: inner hair cells are located innermost along the spiral and outer hair cells are located outermost. There is one row of inner hair cells and usually three rows of outer hair cells. In the human cochlea, there are about 3500 inner hair cells and about 14,000 outer hair cells.

Hair Cells Transduce Mechanical Energy Directly into Electrical Energy

Hair cells (Fig. 25.3) are polarized epithelial cells whose major functions are partitioned into apical and basal cellular compartments. The apical end of the cell is specialized for the reception and translation of mechanical energy into receptor currents, whereas the basal end is specialized for the transmission of information to the central nervous system via synaptic contacts with the primary afferent neurons. Stereocilia are modified microvilli, a fraction of a micrometer in width, that project from the cell apex and contain an abundant supply of tightly packed actin filaments, bound by fimbrin, coursing along their length. Depending on which specific organ the hair cells reside in, their lengths range from a fraction (in the bat cochlea) to tens of micrometers (in vestibular organs). The number of stereocilia per hair cell ranges from about 10 to 300. Together the stereocilia from one cell form a "hair bundle" and are bundled together by extracellular filamentous linkages. They taper rapidly as they insert into the hair cell's cuticular plate, an actin-rich apical cytoplasmic

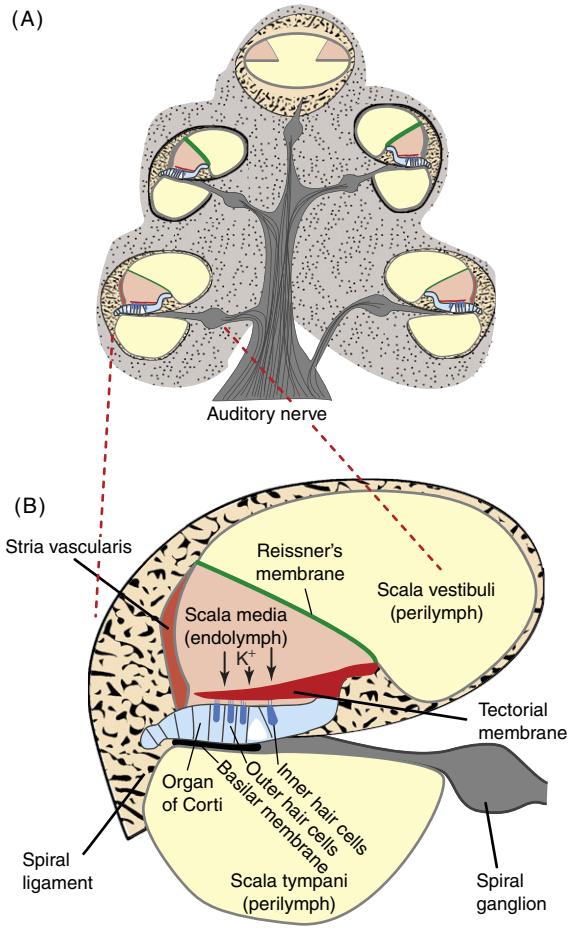


FIGURE 25.2 (A) Cross section through the cochlea. The cross section shows that there are approximately three turns in this human cochlea and that they spiral around a central core (modiolus) that contains the auditory nerve. (B) Cross section through one cochlear turn to illustrate important cell groups (organ of Corti, spiral ligament, stria vascularis, and spiral ganglion) and the main fluid compartments [scala vestibuli, scala media (shaded orange), and scala tympani]. Within the organ of Corti, sensory cells (inner and outer hair cells) are shaded dark blue and are situated between the basilar and tectorial membranes, which move when sound stimulates the cochlea. When these membranes cause motion of the stereocilia of the hair cell, the receptor current, a potassium (K^+) current, flows into the hair cells from the endolymph. The high K^+ concentration and high electrical potential of the endolymph are created by the stria vascularis, and these factors increase the driving force for the receptor current. Supporting cells in the organ of Corti are shaded light blue; these cells form a network that is coupled by gap junctions and may recycle K^+ back to the stria vascularis. Also shown is the spiral ganglion, which contains cell bodies of the auditory nerve fibers.

structure. Because the stereocilia are stiff, rod-like structures, they pivot at their insertion when deflected.

Receptor Potentials Are Evoked by Mechanically Gated Ion Channels within Stereocilia

With the hair bundle in its resting, unperturbed position (Fig. 25.3), a standing inward current exists through a small proportion (10 to 25%) of mechanically activated channels. Because these channels are nonselective for cations, this inward, positively charged flux tends to depolarize the hair cells. In many hair cell sensory systems, including the organ of Corti, the ionic milieu surrounding the hair bundle is richest in potassium (Fig. 25.2B); thus, the major charge carrier for the transduction current is potassium. However, small amounts of calcium are also required to sustain stereociliar channel activity. Displacement of the hair bundle toward the tallest stereocilium increases the proportion of open channels, thereby producing an increase of the inward resting current. The subsequent flow of current across the basolateral membrane produces a depolarizing voltage change, a receptor potential, capable of activating a variety of voltage-dependent conductances in that membrane. Conversely, closure of transduction channels during bundle movement away from the tallest stereociliar row reduces the inward current, effectively hyperpolarizing the basolateral membrane. Several lines of evidence indicate that the flow of transduction current into hair cells occurs near the top of the hair bundle—that is, channels are located near the tips of the stereocilia. Maximal transducer conductance changes up to about 10 nS have been observed. Single channel conductances on the order of 20–200 pS have been either measured or estimated and may increase from apex to base along the tonotopic axis of the cochlea. From such conductance measures, the number of channels per stereocilium has been computed to be one to two.

The relationship between degree of bundle deflection and receptor potential magnitude is neither linear nor symmetric (Fig. 25.4). Bundle displacements in the depolarizing direction produce larger responses than equal displacements in the opposite direction. The displacement–response function is sigmoidal and shifted from its midpoint. Thus, symmetrical sinusoidal deflections of the bundle (as might occur with acoustic stimuli) will produce both sinusoidal (ac) and superimposed depolarizing steady-state (dc) changes in membrane potential. Saturating electrical responses are evoked by deflections as small as 300 nm (Hudspeth & Corey, 1977).

Stereocilia Tip Links May Underlie Transducer Gating

The molecular basis of the hair bundle's response polarity appears to reside in specialized structural attachments at the tips of stereocilia (Pickles, Comis, & Osborne, 1984) (Fig. 25.5). These "tip links" are molecular filaments, comprised of cadhedrin-23 and protocadhedrin 15, that link the top of each stereocilium with a dense membranous plaque on the upper side of the adjacent taller stereocilium, occurring in line with the axis of maximal bundle sensitivity. They are believed to provide the tension required to open transduction channels during bundle deflection, with one end of the filament being anchored and the other pulling on the channel gate located on the shorter stereocilium.

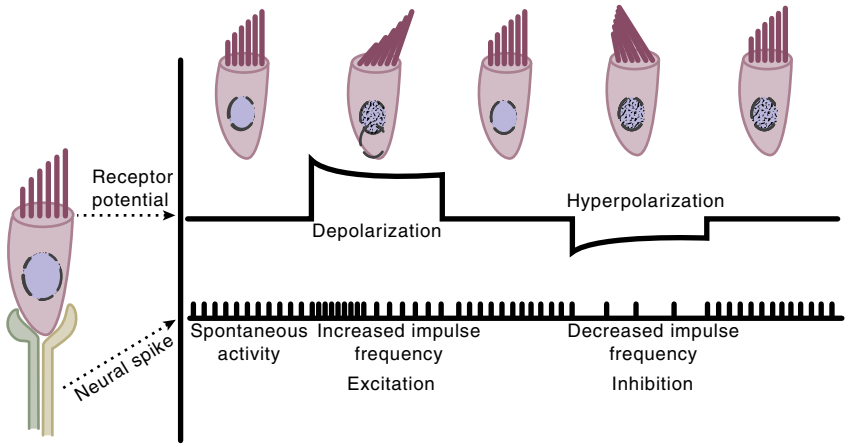


FIGURE 25.3 Illustration of hair cell stimulation, receptor potential response, and auditory nerve fiber discharge. *From Flock (1965).*

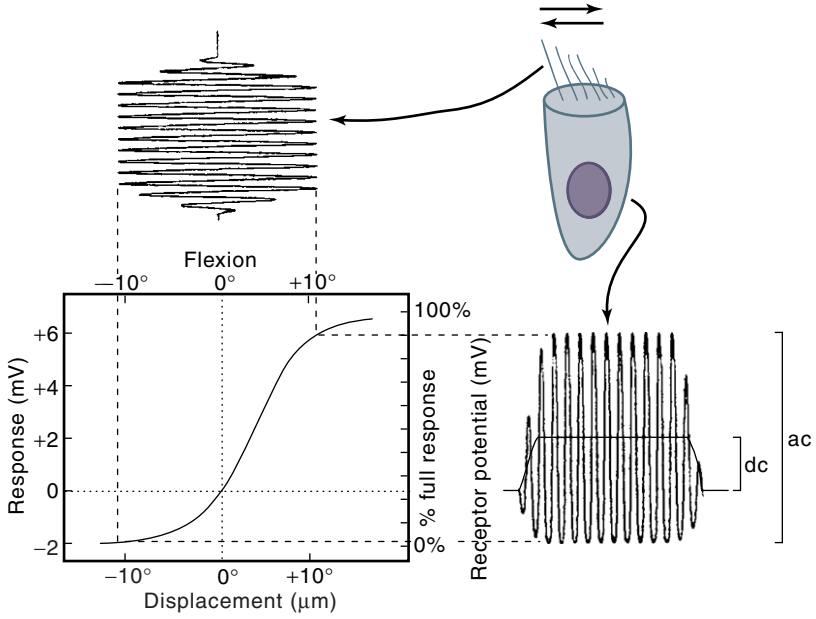


FIGURE 25.4 Sigmoidal input-output function of hair cells. Symmetrical sinusoidal displacement of stereocilia produces an asymmetrical response that includes ac and dc receptor potential components. *Modified from Hudspeth and Corey (1977).*

As the hair bundle tilts during deflection in the excitatory direction, adjacent stereocilia shear against one another, increasing the tension of tip links, thereby increasing the probability that transducer channels will open. When the deflection is in the hyperpolarizing direction, tip link tension slackens and the channels tend to close. In line with tip link or gating spring hypothesis, destruction of the tip links by enzymatic (elastase) or chemical (calcium chelators) treatments can abolish mechanical transduction. In addition, if tip link tension accounts for much of the bundle's stiffness, as expected, bundle compliance varies with the extent of deflection. The deflection versus bundle compliance function is bell shaped, with maximum compliance occurring when half the transduction channels are open. Furthermore, compliance changes are abolished by blocking stereociliar transduction channels with aminoglycoside antibiotics.

Electrical Properties of the Basolateral Membrane Shape Receptor Potentials

The speed of the hair cell transduction process is incredibly fast compared to transduction in other sensory systems, such as vision, olfaction, and the taste modalities of sweet and bitter. The delay between a bundle deflection and the onset of receptor current is estimated to be about 10 μs at 37°C (Corey & Hudspeth, 1983). This rapid response is a consequence of direct mechanical gating of transduction channels, and such speed is essential for auditory hair cells to detect and utilize sound frequencies in the kilohertz (thousand per second) range. Humans can detect sound frequencies up to about 20 kHz, and some mammals, such as bats, can hear above 100 kHz. Although receptor currents may be generated without attenuation

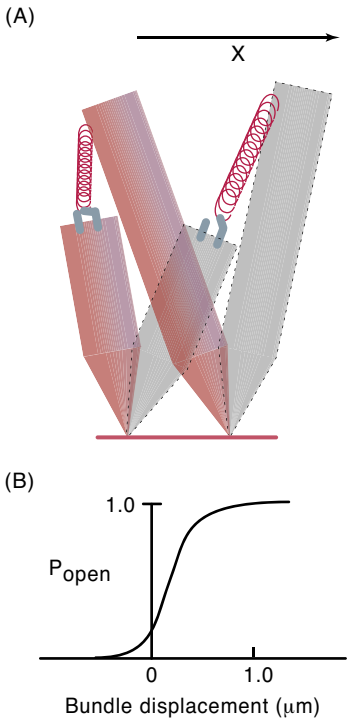


FIGURE 25.5 Gating spring model of hair cell transduction. (A) Tip links connecting channel gate to adjacent taller stereocilium tense during displacement toward the taller stereocilium, thus increasing the probability that the channel will open (B). *From Pickles and Corey (1992).*

across frequency, receptor potentials, which ultimately are responsible for the release of neurotransmitter at the hair cell synapse, are susceptible to the RC filter characteristics of the basolateral membrane. The RC time constant of hair cell membranes ranges from a fraction of a millisecond to a few milliseconds at the resting potential. This translates in the frequency domain to a low-pass filter whose cutoff frequency (fc, the frequency at which the response energy is halved) ranges from tens of hertz to about 1 kHz. This filtering reduces the ac receptor potential in half for every octave increase in frequency above the cutoff. Ultimately, at very high frequencies, ac responses will be negligible, but the dc component will remain unperturbed (Fig. 25.6)—a process termed rectification. The actual cutoff frequency is considered dynamic because receptor potentials themselves may activate voltage-dependent ionic conductances in the basolateral membrane that modify the resistive component of the RC product. In outer hair cells from the organ of Corti, the capacitance of the basolateral membrane is also highly voltage dependent (Santos-Sacchi, 1992), and therefore in this cell type variations in both resistive and capacitive components may influence the membrane filter.

The mammalian organ of Corti rests upon an acellular basilar membrane (Fig. 25.2B) that extends along the coiled cochlea. Cells at the basal end of the organ of Corti have high characteristic frequencies, whereas those at the apex have low ones. Von Bekesy, who won the Nobel Prize in 1968, discovered that different regions of the basilar membrane are tuned to particular frequencies; low-frequency tones cause maximal vibrations of the basilar membrane near the apex, whereas higher-frequency tones cause vibrations more basally. This tuning results from the mechanical characteristics of the basilar membrane, which becomes stiffer and narrower with distance from the apex to the base. As demonstrated in early experiments, the intrinsic frequency selectivity afforded by basilar membrane tuning is not great enough to account for the very selective responses observed in hair cells and auditory nerve fibers. Later experiments demonstrated that active mechanisms are necessary to achieve this high frequency selectivity.

Active Mechanical Properties of the Outer Hair Cells Drive the Mammalian Cochlea Amplifier

Outer hair cells (OHCs) are cylindrically shaped and decrease in resting length by a factor of about four from the apical to the basal ends of the cochlea. When an isolated OHC is stimulated electrically, it responds by altering its length (Brownell, Bader, Bertrand, & de Ribaupierr, 1985). No other auditory cell type responds in this manner. The mechanical response is voltage dependent; depolarizing stimuli induce contractions and hyperpolarizing stimuli induce elongations, with responses as large as 30 nm/mV (Fig. 25.7). This mechanical activity is not akin to any other known form of cellular motility and is governed directly by voltage-dependent, integral membrane protein motors. The motor is the gene product of Prestin (Zheng et al., 2000), one of a family of anion transporter genes (SLC26). OHC electromotility has been demonstrated, *in vitro,* to extend well into the tens of kiloHertz range. *In vivo*, the motor action of OHCs probably is responsible for otoacoustic emissions (Box 25.1).

Through a variety of experimental approaches, OHC motor activity has been shown to be restricted to the lateral membrane of the OHC. As might be expected for a voltage-dependent process that resides within the membrane, a charged voltage sensor must exist, just as voltage-dependent ion channels have voltage sensors. The existence of an OHC motility voltage sensor is confirmed by measuring gating charge movements (capacitive-like currents) under voltage clamp while blocking ionic conductances. These gating currents represent the restricted movement of the charged voltage sensor within the plane of the lateral membrane; increasing voltage will move more charge (thus activating more motors) according to Boltzmann statistics (Fig. 25.8A). Maximum charge moved is about 7500 e/μm^2, and may represent the density of motor molecules in the lateral plasma membrane.

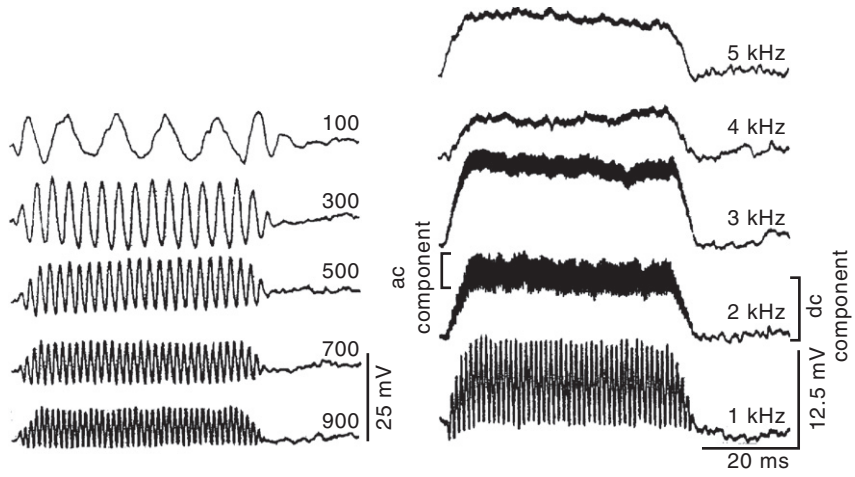


FIGURE 25.6 Receptor potentials from the mammalian inner hair cell in response to acoustic bursts of increasing frequency. Because of the cell's *RC* time constant, the ac component diminishes as frequency increases. However, the dc component remains intact. *From Russell and Sellick (1983).*

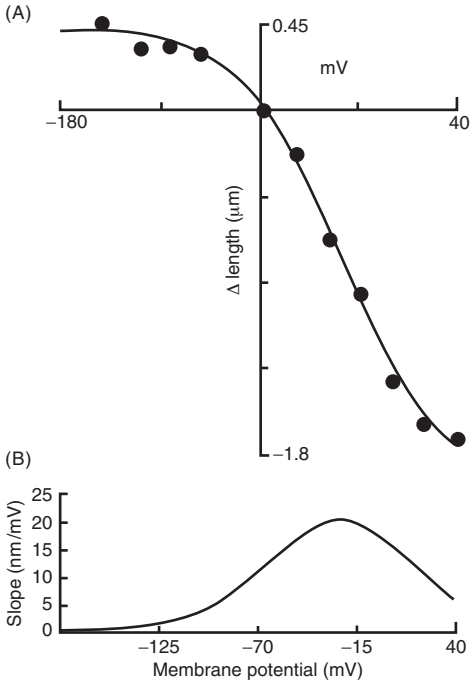


FIGURE 25.7 Mechanical response of mammalian outer hair cell (OHC) under voltage clamp. The OHC changes its length when the cell is held at different membrane potentials (A). The slope of the sigmoidal input-output function defines the cell's sensitivity to membrane potential change (B). *From Santos-Sacchi (1992).*

The slope of the Q/V function is defined as capacitance, and thus the capacitance of the OHC is a bell-shaped function of voltage (Fig. 25.8B). This nonlinear capacitance rides atop the cell's intrinsic linear membrane capacitance of 1 μF/cm^2. Thus, in an OHC of about 70 μm in length, at the point where motile gain is maximum (i.e., where half the motors are activated), the capacitance peaks at about double the cell's linear capacitance. As mentioned earlier, such nonlinear capacitance may have significant effects on membrane-filtering characteristics. Although the mechanical response is voltage dependent and is not evoked by activation of any particular voltage-dependent ionic conductance, the activity of the voltage sensor requires intracellular chloride (Rybalchenko & Santos-Sacchi, 2003), and in intact OHCs a stretch-activated chloride conductance appears crucial for maintaining and modulating motor activity.

Contributions of lateral membrane activity and stereocilia activity, each of which is capable of force production, are likely involved in cochlear amplification. Active mechanical responses of the stereocilia bundle may be driven by calcium influx through calcium-sensitive, mechanically activated stereocilia channels. Indeed, recent evidence suggests that bundle activity, acting through its high pass filter characteristic, can provide for a tuned band pass response of electromotility.

To summarize, OHCs, probably through a mechanical feedback scheme, are required to boost basilar membrane motion and enhance frequency selectivity, a process termed the "cochlear amplifier." In the absence of OHCs or with an impairment of prestin function, either by gene mutation (Dallos et al., 2008) or by reducing OHC chloride levels *in vivo* (Santos-Sacchi, Song, Zheng, & Nuttall, 2006), hearing sensitivity and frequency selectivity are impaired. OHCs are unique because they function as both receptors and effectors, transducing mechanical stimuli via hair bundle displacement and changing length and other mechanical properties in response to the generated receptor potentials. These mechanical events, via an energetic boost to basilar membrane motion, provide an enhanced stimulus to the inner hair cells, which are the cells that provide synapses to the overwhelming majority of auditory nerve afferents.

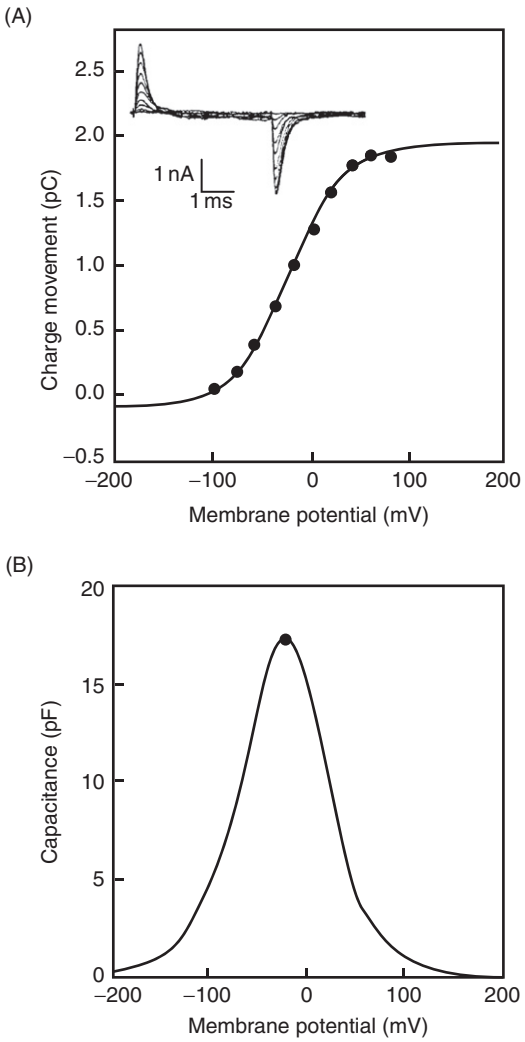


FIGURE 25.8 Gating charge associated with OHC motility voltage sensor. When the membrane potential is stepped with increasingly larger depolarizing voltages from a negative holding potential, nonlinear capacitive currents are generated and are obvious after linear capacitive currents are subtracted (inset). Integrating the onset currents, a measure of the amount of charge moved within the membrane can be obtained (A). The function is sigmoidal and has characteristics similar to the mechanical response. The first derivative of the charge with respect to membrane voltage defines the cell's nonlinear capacitance (B). *From Santos-Sacchi (1991).*

Cochlear Endolymph Increases the Sensitivity of Hearing

An increase in the sensitivity of hair cells is made possible by the unique composition of the inner ear fluids. Within the inner ear, fluids are contained in three compartments known as scalae: scala tympani, scala media, and scala vestibuli (Fig. 25.2B). These scalae extend in parallel along the length of the cochlea from the base to the apex. Scala tympani and scala vestibuli contain perilymph, which is high in Na^+ and low in K^+, similar to other extracellular fluids. Scala media contains endolymph, a specialized fluid with a low concentration of Na^+ and a high concentration of K^+ (about 160 mM). The endolymph is also unusual because it has a positive electrical potential (about 90 mV), observable even in the absence of sound. The endolymph is generated by a highly vascularized tissue in the lateral edge of the cochlea, the stria vascularis (Fig. 25.2B).

The organ of Corti, which contains the hair cells, is located at the junction between endolymph and perilymph (Fig. 25.2B). Stereocilia are surrounded by endolymph, whereas the remainder of the hair cell (basolateral membrane) is surrounded by perilymph. The endolymph increases the sensitivity of the hair cells because it increases the transduction current, a K^+ current that flows from the endolymph into the hair cells (Fig. 25.2B). The endolymph increases this current because (1) the high concentration of K^+ in endolymph forms a concentration gradient that establishes a driving force for K^+ flow into the cells and (2) the positive potential of the endolymph forms a large electrical gradient that is also a driving force for K^+ flow into the hair cells. Once the chemical and electrical properties of the endolymph are established and sound stimulates the cochlea, K^+ flows into the hair cells with little energy expenditure by the hair cells, as K^+ is flowing down its electrochemical gradient. Thus, hair cells do not require a high blood flow that might bring noise to the organ of Corti, interfering with sound reception. In fact, there are only a few blood vessels near the organ of Corti.

Sensorineural Hearing Loss Often Results from Damage to Hair Cells

Sensorineural hearing loss results from damage to hair cells or, less commonly, from damage to auditory nerve fibers (reviewed by Merchant & Nadol, 2010). Hair cells can be destroyed or their hair bundles damaged by intense sounds such as those generated by guns, jet engines, or even earphones operated at high sound levels. Hair cell loss is permanent in the mammalian cochlea, but in the bird cochlea, hair cells are regenerated from nearby supporting cells (Corwin & Cotanche, 1988). Damage from intense sounds can be minimized by (1) decreasing the duration of the sound exposure and (2) decreasing the level of the sound at the tympanic membrane (by wearing protectors such as earguards and earplugs). Hair cells can also be destroyed by chemical agents like aminoglycoside antibiotics (e.g., streptomycin, kanamycin). These agents block the transduction channels directly, but their mechanism of hair cell destruction is independent. In individuals with sensorineural hearing loss, some hearing can be restored with a cochlear implant (Box 25.2).

Summary

Current research on the cochlea is centered on the mechanisms of outer hair cell motility and how this motility

BOX 25.1

OTOACOUSTIC EMISSIONS

In 1978, David Kemp made a surprising discovery: the ear can sometimes *emit* sounds. These "otoacoustic emissions" are now the focus of much interest because basic researchers can use them to study the function of the ear and because clinicians can use them as tests of hearing (reviewed by Shera, 2004). Emissions are almost always so low in level that they are inaudible unless the individual is in an exceptionally quiet environment. Thus, measurement of emissions requires a sensitive, low-noise microphone that is placed in the external ear canal. Many studies indicate that the cochlea (inner ear) is the source of emissions. Within the cochlea, the outer hair cells are likely to be the generators of emissions. After emissions are generated, they travel in reverse of the normal pathway for sound into the inner ear: the emissions propagate from their point of generation along the basilar membrane to the oval window, then through the middle ear via the ossicles, and finally they move the tympanic membrane to result in airborne sound.

There are two main types of emissions—spontaneous and evoked—the latter are evoked in response to an externally presented sound. Spontaneous emissions are detected from the ears of about one-half of normal hearing humans, but these emissions are rare in laboratory animals. They are almost always pure tones. Transient-evoked emissions are evoked by a short sound such as a click and appear several to tens of milliseconds later. These emissions were first called the "cochlear echo," but they are not a real echo because more energy can appear in the emission than was present in the evoking sound. Distortion product-evoked emissions (Fig. 25.9) are evoked by two tones ("primaries" of frequencies f_1 and f_2) and occur at combinations of the primary frequencies (such as at frequency $2f_1–f_2$). The two primary tones each produce traveling waves along the basilar membrane, and a likely point of generation of the emission is where these waves overlap maximally.

Otoacoustic emissions can be used clinically as a screening tool in hearing tests, as subjects with sensory hearing losses greater that 30 dB typically lack emissions. Transient-evoked or distortion product-evoked emissions are used in such tests. Emission-based tests are especially valuable for individuals such as infants who are otherwise hard to test by conventional audiometry. Otoacoustic emissions only rarely are the cause of tinnitus, the sensation of ringing in one's ears. Most individuals who have tinnitus do not have emissions corresponding to the tinnitus. Thus, tinnitus arises by some other mechanism, almost certainly within the nervous system.

M. Christian Brown

Reference

Shera, C. A. (2004). Mechanisms of mammalian otoacoustic emission and their implications for the clinical utility of otoacoustic emissions. *Ear and Hearing, 25,* 86–97.

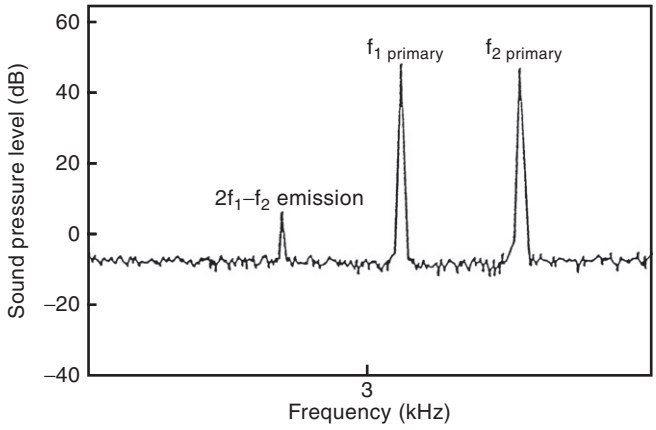


FIGURE 25.9 Distortion-product otoacoustic emission from a human subject measured with a microphone placed in the ear canal. The microphone records the sound pressure of the two primary tones (frequencies f_1 = 3.164 kHz and f_2 = 3.828 kHz, each 50 dB sound pressure level that were used to evoke the emission at $2f_1–f_2$ (2.5 kHz at 12 dB). *From Lonsbury-Martin and Martin (1990).*

contributes to basilar membrane motion. The mechanisms of how sound might use active stereocilia motion remain to be worked out. Substantial interest focuses on the processes and factors involved in hair cell regeneration with the eventual hope that approaches like stem cell implants might encourage restoration of function in the mammalian cochlea. The improvement of cochlear implants is also a very active area of clinical research.

BOX 25.2

COCHLEAR IMPLANTS

The cochlear implant is one of the most successful prostheses used to stimulate the nervous system (Niparko et al., 2000). The cochlear implant can provide partial restoration of hearing for individuals with sensorineural hearing loss. In sensorineural hearing loss, there is usually a partial or complete loss of the sensory cells (hair cells) in the inner ear (cochlea). Hair cells normally transduce the mechanical energy of sound into the electrical energy of receptor potentials. They synapse on primary auditory nerve fibers, which send information to the brain. Hair cells can be damaged irreversibly by intense sound, ototoxic drugs, or the aging process. Once lost, hair cells are not regenerated in mammals. Often, however, individuals with hair cell loss retain a significant complement of auditory nerve fibers. It is these fibers that are stimulated by the cochlear implant.

The cochlear implant consists of a microphone to detect sound, an electronic "processor" that transforms the sound waveform into a code of electrical stimuli, and an array of stimulating electrodes in the cochlea. The electrode array (Fig. 25.10) is inserted through or near the round window into scala tympani where it lies close to the peripheral axons of primary auditory neurons. Usually, the implant consists of more than 20 electrodes that begin in the base and are spaced at intervals along the cochlear spiral. Those electrodes in the most basal regions are positioned to stimulate nerve fibers that originally responded to high frequencies and those in more apical regions are positioned to stimulate fibers that originally responded to lower frequencies.

In some individuals who have received implants, comprehension of speech is possible, even when there are no other cues such as lip reading. These individuals can carry on a normal conversation, even via telephone. In other individuals, however, full speech comprehension is not restored but the implant, together with lip-reading ability, assists with spoken conversation. It also provides for the detection of important sounds such as the ring of a telephone and the approach of a vehicle. Individuals who have become deaf after the acquisition of spoken language have better comprehension than prelingually deafened individuals, almost certainly because of differences in the central nervous system.

M. Christian Brown

Reference

Niparko, J. K., Kirk, K. I., Mellon, N. K., Robbins, A. M., Tucci, D. L., & Wilson, B. S. (Eds.), (2000). *Cochlear implants: Principles and practices*. Philadelphia, PA: Lippincott Williams & Wilkins.

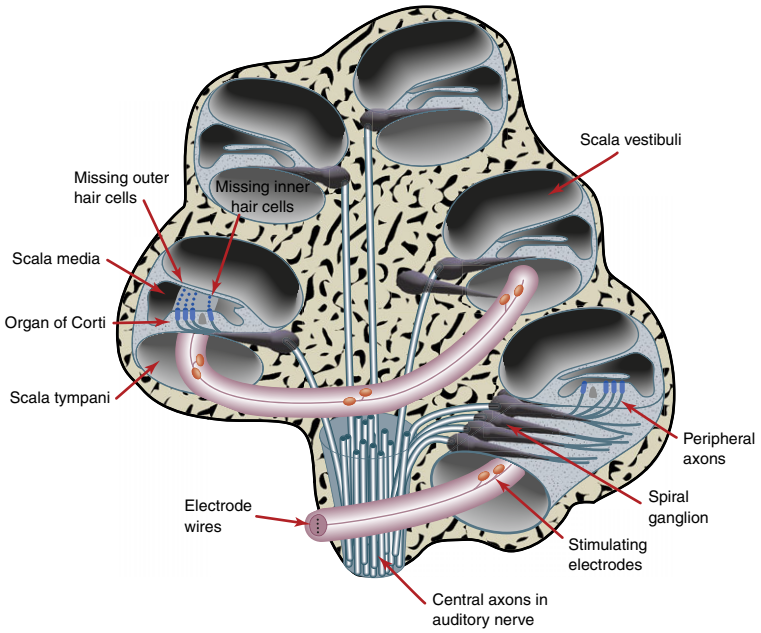


FIGURE 25.10 Drawing of the human cochlea showing cochlear implant electrodes. The electrode array is inserted through the round window of the cochlea into the fluid-filled space called the scala tympani. It stimulates peripheral axons of the spiral ganglion neurons, whose central axons (auditory nerve fibers) send impulses into the brain. In the normal cochlea, frequency is mapped along the cochlear spiral, with the lowest frequencies at the apex of the cochlea (at the top of the figure). The different electrodes of the cochlear implant are designed to stimulate different groups of nerve fibers that originally responded to different frequencies, although because of spatial constraints the implant is not inserted all the way to the cochlear apex. *From Loeb (1985).*

THE AUDITORY NERVE

Hair cells receive their primary afferent innervation from neurons of the spiral ganglion, located in the central core, or modiolus, of the cochlea (Figs. 25.2 and 25.10). These bipolar auditory neurons send peripheral axons to the hair cells and central axons into the brain by way of the auditory nerve, a subdivision of the eighth cranial nerve. Two types of afferent neurons separately innervate the inner and outer hair cells (Fig. 25.11). The first type (the type I neuron) sends processes to contact inner hair cells, almost always contacting a single hair cell, whereas the second type (the type II neuron) sends processes to contact from 5 to 100 outer hair cells. Fibers from type I neurons are relatively large in diameter and myelinated; thus, their information reaches the brain quickly, within a few tenths of a millisecond. Fibers from type II neurons are thin and unmyelinated and transmit information much more slowly. Both types of afferent fibers project centrally into the cochlear nucleus in the brainstem. Interestingly, type I neurons total about 95% of the afferent population (about 30,000 in humans), whereas type II neurons total only about 5%. Thus, outer hair cells, which number over three-quarters of the receptor cell population, are innervated by only a small minority of the afferent neurons. This innervation plan strongly suggests that the functional role for inner hair cells and type I neurons is to serve as the main channel for sound-evoked information flow into the brain.

Responses Are Sharply Tuned to Frequency

Type I auditory nerve fibers respond to sound and transmit these responses to the brain via discrete action potentials (reviewed by Ruggero, 1992; Geisler, 1998). The brain must extract information from these spikes and process this information to eventually form a percept of the stimulus. The information available to the brain via the auditory nerve consists of which nerve fibers are responding and the rate and time pattern of the spikes in each fiber. A fiber's response area is usually plotted as a graph of sound pressure level (SPL) versus frequency with the line showing the threshold contour; such a graph is known as a tuning curve (Fig. 25.12). The lowest point on the tuning curve corresponds to the characteristic frequency (CF). At CF, auditory nerve fibers can respond to sound levels as low as 0 dB in the most sensitive range of hearing. At low sound levels, the tuning curve is impressively narrow, indicating that the fiber responds only to a narrow band of frequencies near CF. This sharply tuned "tip" region (Fig. 25.12) is likely generated by the active motility of the outer hair cells. At high sound levels, the tuning curve becomes much wider, especially for frequencies below CF. This response to a broad range of frequencies likely reflects the passive mechanical characteristics of basilar membrane motion with little contribution from outer hair cells.

The relationship between CF and its point of innervation in a type I fiber has been demonstrated by single unit labeling (Liberman, 1982). The distance to the point of innervation along the length of the cochlea is logarithmically related to fiber CF. Fibers with the lowest CFs innervate the apex of the cochlea, and fibers with progressively higher CFs innervate progressively more basal positions, as expected from the pattern of basilar membrane vibration. This precise mapping of frequency to position is known as tonotopic mapping, and it strongly suggests that frequency is coded via a "place" code.

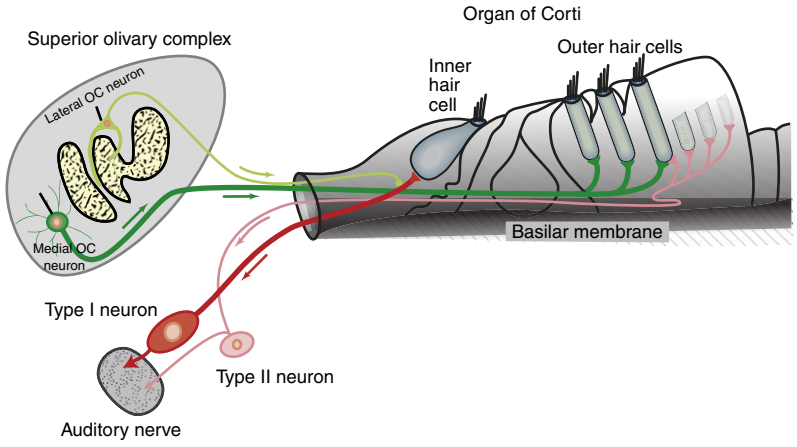


FIGURE 25.11 Innervation patterns of afferent and efferent neurons in the organ of Corti. Afferent innervation is provided by spiral ganglion neurons, which have central axons that form the auditory nerve. There are two types of afferent neurons: (1) type I neurons, which receive synapses from inner hair cells, and (2) type II neurons, which receive synapses from outer hair cells. Efferent innervation is provided by a subgroup of neurons in the superior olivary complex that send axons to the cochlea and are hence called olivocochlear (OC) neurons. There are two types of OC neurons: (1) lateral OC neurons, which innervate type I dendrites near inner hair cells, and (2) medial OC neurons, which innervate outer hair cells. Lateral OC neurons are distributed mainly ipsilateral to the innervated cochlea, whereas medial OC neurons are distributed bilaterally, with approximately two-thirds from the contralateral side (not illustrated) and one-third from the ipsilateral side of the brain. *From Warr et al. (1986).*

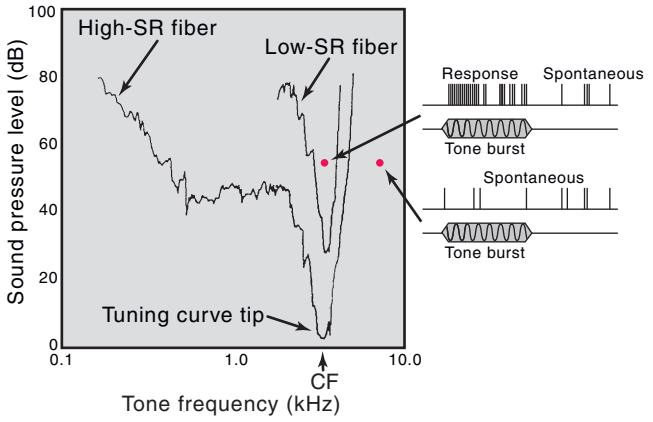


FIGURE 25.12 Tuning curves of two type I auditory nerve fibers. These curves plot the sound pressure level necessary to cause a response as a function of sound frequency. Within the tuning curve, the fiber responds to sound, whereas outside the tuning curve, there is only spontaneous firing (insets at right). The frequency corresponding to the lowest point on the tuning curve is called the characteristic frequency (CF); it is the point of maximal sensitivity. The sharply tuned region near the CF is called the tuning curve tip. The sharp tuning and high sensitivity of this region are generated by the mechanical properties of the outer hair cells. Tuning curves from two fibers are shown, one from a high SR (spontaneous rate) fiber and another from a low SR fiber. As is typical, the high SR fiber has the highest sensitivity. *From Kiang (1984).*

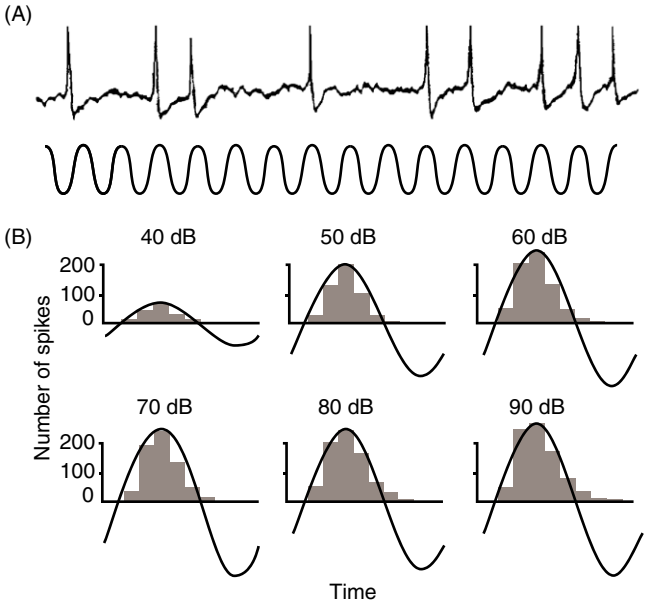


FIGURE 25.13 (A) Preferential firing of an auditory nerve fiber at a certain phase of the sound waveform. This pattern is called "phase locked," although the firing is not on every cycle of the waveform. Stimulus frequency was 0.3 kHz. (B) Histograms that quantify the time of firing plotted within one cycle of the sound waveform for many repeated cycles of the stimulus. Response of the fiber is phase locked at the moderate and high SPLs shown, even on this very fast time scale (time for one period was about 0.9 ms, given the stimulus frequency of 1.1 kHz). (A) *From Evans (1975)*; (B) *From Rose et al. (1971).*

Phase Locking of Responses and Codes for Sound Frequency

Responses of auditory nerve fibers can show time-locked discharges at particular phases within the cycle of the sound waveform, a property known as phase locking (Fig. 25.13). Although locked to a particular phase, there is generally not a spike for every waveform peak. Phase locking in nerve fibers results from the phasic release of neurotransmitter as dictated by the ac receptor potential (Fig. 25.6). Phase locking and the ac receptor potential decrease for frequencies above 1 kHz. At low sound frequencies, phase locking is an important temporal code for sound frequency and the sensation of pitch.

Neural Response Is a Function of Sound Level and Spontaneous Activity

The response of a single auditory nerve fiber increases with sound level until a point at which the rate of the fiber no longer increases and is saturated (Fig. 12.14A, solid line). The dynamic range over which the rate of most fibers increases is generally between 20 and 30 dB. How then can the auditory nerve signal the large range in level of audible sound from 0 to 100 dB? First, it is likely that as the level of a tone increases, more and more fibers that are tuned to other CFs begin to respond, because tuning curves become broader at higher sound levels (Fig. 25.12). Second, auditory nerve fibers vary in their sensitivity to sound, and as sound level is increased, the less sensitive fibers begin to respond. Sensitivity of fibers at a given CF varies by as much as 70 dB. The sensitivity of response is correlated with the rate of spontaneous firing, which is the rate of firing when there is no stimulus or when the stimulus is outside the tuning curve (Fig. 25.12). Spontaneous rates (SRs) vary from one fiber to another over the range of 0 to 100 spikes/s. Although there may be a continuum of SR, three main groups of fibers have been defined (low SR: <0.5 spikes/s; medium SR: 0.5 to 17.5 spikes/s; high SR: >17.5 spikes/s), and these groups predict many physiological and anatomical characteristics of auditory nerve fibers. The high SR fibers have higher sensitivities than medium and low SR fibers (Fig. 25.12). Low and medium SR fibers give off the largest number of terminals in the cochlear nucleus of the brainstem and preferentially innervate certain regions, such as the peripheral cap of small cells, suggesting that information carried by the different SR groups may be processed in somewhat different regions. Low SR fibers may be less sensitive, but they likely play important roles in detecting changes in sounds at high sound levels. Low SR fibers can signal changes at high sound levels because their low sensitivity causes them to respond mostly at higher sound levels and because there is less tendency for their firing rates to saturate.

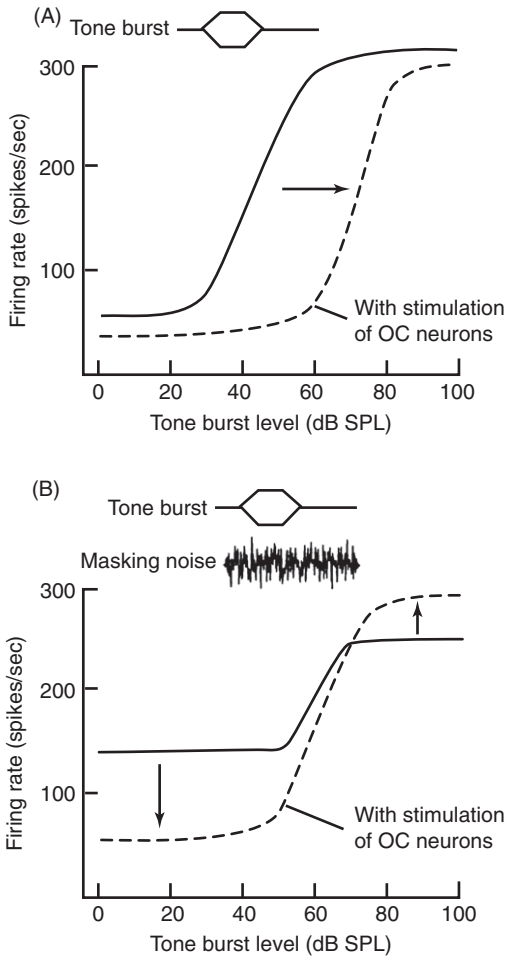


FIGURE 25.14 Rate level function for an auditory nerve fiber in response to tone bursts without (solid lines) and with (dashed lines) electrical stimulation of olivocochlear (OC) neurons. (A) With tone bursts alone, the discharge rate rises with sound level until it reaches a maximum and no longer increases (saturation). Stimulation of the OC neurons shifts the function to the right toward higher tone burst levels (arrow). This shift adjusts the dynamic range of the fiber so that it can signal changes in tone burst level for higher sound levels; this is likely to be an important function of OC neurons. (B) Level function from the same fiber, now for tone bursts accompanied by continuous masking noise (insets at top). At low levels of tone bursts, the fiber has a significant firing rate because it is responding mainly to the noise. Even at high levels of tone bursts, the fiber is still responding to the noise in between tone bursts, which causes adaptation, and the fiber responds less to the tone burst than with tone bursts alone. In this case, stimulation of OC neurons (dashed line) decreases the response to the noise, thus decreasing the rate of the fiber at low levels of tone bursts (left arrow). Because the fiber is responding less to the noise, it is less adapted and has a greater response at high levels of the tone burst (right arrow). The fiber now has a greater ability to signal changes in level of the tone burst; this effect has been called "antimasking" and is likely to be another important function of the OC system. *Adapted from Winslow and Sachs (1987).*

Masking Involves Adaptation and Suppression

The auditory nerve response to one stimulus may be masked, or hidden, by the presence of other stimuli. Masking involves properties of auditory nerve response such as adaptation and two-tone suppression, and possibly refractory properties. In adaptation, firing to a tone burst is high initially and then lessens, or adapts, to a steady state over time (Fig. 25.15). Adaptation probably takes place at the hair cell/nerve fiber synapse, as there is little adaptation in receptor potentials of hair cells for these short times (Fig. 25.6). Because of adaptation to one stimulus, the fiber is less likely to respond to a second stimulus; that is, the first stimulus masks the second stimulus. Masking by continuous noise changes the responses to tone bursts greatly. With tone bursts alone (Fig. 25.14A, solid line), there is a moderate dynamic range and a large difference in firing rate from low to high tone burst levels. With tone bursts in masking noise (Fig. 25.14B, solid line), there is a substantial rate at low tone burst levels because the fiber is now responding to the masking noise. At high tone burst levels, the rate is decreased because the fiber is adapted by the noise and is less likely to respond to the tone burst. Thus, with masking noise, there is much less difference in the rate of the fiber as a function of tone burst level as well as a lower dynamic range: these effects decrease the ability of the fiber to signal changes in the tone burst level.

An additional type of masking, referred to as suppressive masking, involves a phenomenon called two-tone suppression. In two-tone suppression, one tone lowers the response to a second tone even though the first tone may not excite the auditory nerve fiber. Two-tone suppression is present in the motion of the basilar membrane. It may help control the amount of gain provided by the cochlear amplifier, because as stimuli composed of several frequencies increase in sound level, two-tone suppression decreases the response of the nerve fiber and extends its dynamic range. Whether there is adaptive masking or suppressive masking depends on the frequencies and levels of signals relative to the response area of the fiber. Masking is important in the auditory system because many auditory processes, such as speech comprehension, are affected greatly by maskers and because background or masking signals are common in many everyday situations.

Descending Systems to the Hair Cells and Nerve Fibers

The auditory periphery and other hair cell organs receive an efferent innervation from the brainstem (Fig. 25.16). The cochlear efferent neurons have cell bodies in the superior olivary complex and project to the cochlea and hence are called olivocochlear (OC) neurons (reviewed by Ryugo, Popper, & Fay, 2011). There are two groups of OC neurons—medial OC and lateral OC neurons (Fig. 25.11)—named according to the positions of their cell bodies in the superior olive. These olivocochlear fibers

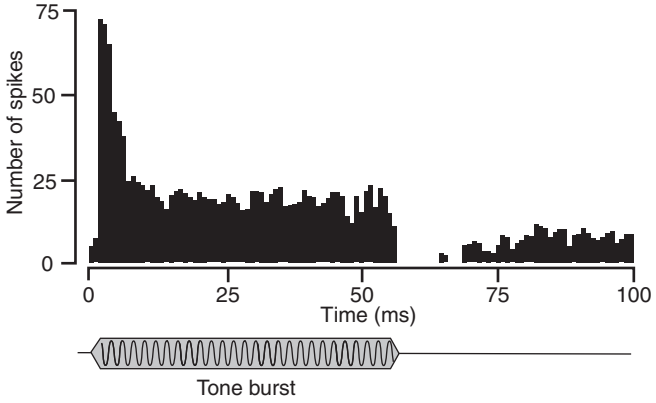


FIGURE 25.15 Poststimulus time (PST) histogram from an auditory nerve fiber in response to a tone burst (outline indicated below). A PST histogram is constructed by repeatedly presenting a stimulus while counting the number of action potentials that fall into bins of time during and after the stimulus. The histogram can be thought of as the probability of firing as a function of time. This function has an initial peak and then a decrease in firing (adaptation) during the burst. After the burst ends, spontaneous firing is lessened but then returns gradually.

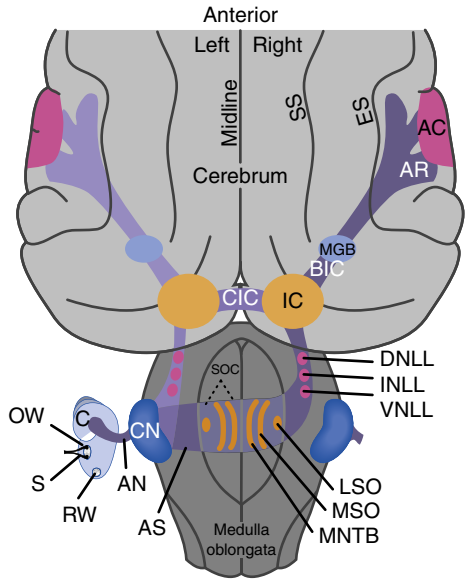


FIGURE 25.16 Simplified schematic of the pathways of the ascending auditory system of a generalized mammal. The pathway begins at the lower left with the cochlea and ends at top right with the auditory cortex. The large band crossing the midline indicates the crossing of the bulk of the pathway below the inferior colliculus. Some ipsilateral pathways and frequent commissural pathways explain the responses of some units at higher centers to ipsilateral sound. AC, auditory cortex; AN, auditory nerve; AR, auditory radiations; BIC, brachium of the inferior colliculus; C, cochlea; CIC, commisure of the inferior colliculus; CN, cochlear nucleus; DNLL, dorsal nucleus of the lateral lemniscus; ES, ectosylvian sulcus; IC, inferior colliculus; INLL, intermediate nucleus of the lateral lemniscus; LSO, lateral superior olive; MGB, medial geniculate body; MNTB, medial nucleus of the trapezoid body; MSO, medial superior olive; OW, oval window; RW, round window; S, stapes; SOC, superior olivary complex; SS, suprasylvian sulcus; VNLL, ventral nucleus of the lateral lemniscus. *From Kiang and Peake (1988).*

separately innervate the two types of hair cells. Medial OC neurons innervate outer hair cells, whereas lateral OC neurons innervate the inner hair cell region by synapsing on dendrites of type I auditory nerve fibers. Little is known about the lateral OC neurons, as their very thin axons are difficult to study. Medial OC neurons can be activated experimentally by electrical stimulation, which causes them to release the neurotransmitter acetylcholine. At the outer hair cell, acetylcholine acts on a nicotinic receptor that allows Ca^{2+} influx, which then opens Ca^{2+}-activated K^+ channels, allowing K^+ efflux that hyperpolarizes the cell. The hyperpolarization and accompanying decrease in input resistance of the cell probably reduces the electromotility of the outer hair cell, decreases basilar membrane motion, and reduces the responses of inner hair cells and auditory nerve fibers. These decreases shift responses of auditory nerve fibers to higher sound levels (Fig. 25.14A, dashed line). This shift means that sound levels that previously saturated the discharge rates are now within the increasing portion of the rate level curve of the fiber, so the fiber can now signal changes in sound level at higher sound levels. Thus, one function of medial OC neurons may be to control the gain of the cochlear amplifier to shift the dynamic range of hearing.

Another function of medial OC neurons may be to reduce the effects of masking noise on the responses of auditory nerve fibers. Decreases in responses by maskers can occur by adaptation and suppression. The effect of OC stimulation on masked responses to tone bursts can be to *enhance* the response (Fig. 25.14B, dashed line). Responses at low tone burst levels are decreased because the response of the fiber to the noise is decreased (left arrow on Fig. 25.14B); responses at high tone burst levels are increased because there is less noise-induced adaptation and a greater response to the tone bursts (right arrow on Fig. 25.14B). The fiber now has a greater ability to signal changes in level of the tone burst; this effect has been called "antimasking" and may be an important function of the OC system. A third function of medial OC neurons may be to protect hair cells in the cochlea from damage due to intense sounds via their large synaptic endings on outer hair cells. Additional protection from damage is afforded by two muscles of the middle ear, whose contractions decrease sound transmission through the middle ear.

Summary

Although the function of type I auditory nerve fibers is fairly well understood, we do not yet know how afferent fibers of the outer hair cells, the type II fibers, function in the hearing process. Future experiments will provide insight into the roles played by different spontaneous rate groups of type I auditory nerve fibers. More work is

needed to identify the mechanisms used to prevent the degradation of signals by masking noise. These mechanisms could potentially be applied clinically in the design of hearing aids and cochlear implants.

CENTRAL NERVOUS SYSTEM

Auditory Pathways Are Tonotopically Organized

The auditory nerve projects tonotopically into the cochlear nucleus, which in turn projects to the other auditory nuclei of the brainstem: the superior olivary complex, nuclei of the lateral lemniscus, and inferior colliculus (Fig. 25.16). The finding of tonotopic mappings in central auditory nuclei supports the theory of a place code for sound frequency (reviewed by Ehret & Romand, 1997; Popper & Fay, 1992; Webster et al., 1992). This is especially true in higher nuclei because here neurons have a decreased ability to phase lock to the sound waveform, and thus a temporal code does not seem to be as robust. In most auditory nuclei there are not large overrepresentations of certain frequencies, although, as noted later, a given nucleus may be more devoted to low frequencies (e.g., medial superior olive) or to high frequencies (e.g., lateral superior olive). An important exception is the case of echolocating bats that use a constant frequency in their echolocating pulse; in these species, there is an over-representation of one of the constant frequencies. This over-representation begins in the cochlea and is preserved in the cortex (e.g., Fig. 25.23C). In these bats much cortical area is devoted to the constant frequency, analogous to the large somatosensory cortical areas devoted to important areas such as the hands and face.

Under certain conditions, tonotopic mappings have been shown to be capable of plastic changes (Robertson & Irvine, 1989; reviewed by Parks, Rubel, & Fay, 2004). After peripheral hearing loss, tonotopic mapping of the auditory cortex is altered such that a region that originally processed frequencies within the hearing loss begins to respond to adjacent frequencies. For instance, when the cochlea is damaged so that it no longer responds to high frequencies, the high frequency portion of the cortex does not stay unresponsive, but over time becomes responsive to middle frequencies where hearing is still normal. Plasticity is of interest to humans because of a common condition called presbycusis, which is hearing loss with advanced age. Presbycusis generally begins with a loss at high frequencies and spreads to lower frequencies with advancing age. While presbycusis and other forms of hearing loss are caused by peripheral changes, they may result in plastic changes in the central pathways that change the way sound is processed in the brain.

Units Are Classified by Their Responses to Sound

Classification by PST Histograms

The cochlear nucleus is the auditory center that is understood best at the cellular level (reviewed by Rhode & Greenberg, 1992). Cochlear nucleus neurons have been well classified both anatomically and physiologically, and structure–function correlations can be made between these classifications. Physiologically, cochlear nucleus neurons process their inputs from the auditory nerve to form new output patterns. These patterns may be used to classify units on the basis of the shape of the poststimulus time (PST) histogram, which plots the spike pattern of the unit as a function of time for short-duration tone bursts. Unit types are called "pauser," "onset," "primary-like with notch," "chopper," and "primary-like" (Fig. 25.17). A correspondence has been established between these unit types and anatomical cell types of the cochlear nucleus. The correspondence was first suggested by the regional distributions of unit and cell types; for example, there is a region of the cochlear nucleus in which one finds mostly "octopus" cells and from which mostly onset units are recorded. Direct correspondence has been made by single-unit labeling, in which a single unit is first classified physiologically according to PST type and subsequently injected with a neural tracer (e.g., horseradish peroxidase) that fills the neuron and its processes. The anatomical cell type can then be determined from postexperiment histology. These types of experiments are difficult and low yielding, but they firmly establish the structure–function correspondences, at least for neurons large enough to record and label *in vivo.* For instance, a major cell type in the ventral subdivision of the cochlear nucleus, the "spherical bushy" cell, has a "primary-like" response pattern (Fig. 25.17).

Use of PST histograms is helpful not only in classifying units but also reveals functional properties of the neurons. Let us consider the example of primary-like units. The primary-like pattern reveals that the spherical bushy cell has faithfully preserved the temporal properties of the auditory nerve fiber response and implies that these units receive endbulbs from nerve fibers (see Box 25.3). In contrast, other types of cochlear nucleus neurons alter the spike patterns of the auditory nerve fibers and presumably have distinct functions. For instance, onset units respond mainly at the onset of a short tone burst (Fig. 25.17). One subclass of onset units can increase its response over a large range of SPL and might be involved in signaling sound level. Relative to mammals, tests of such hypotheses are easier in birds, where there is much more regional segregation of neurons by anatomical and physiological type in the cochlear nucleus. In birds, bushy cells are segregated in one region (nucleus magnocellularis), whereas other cell types,

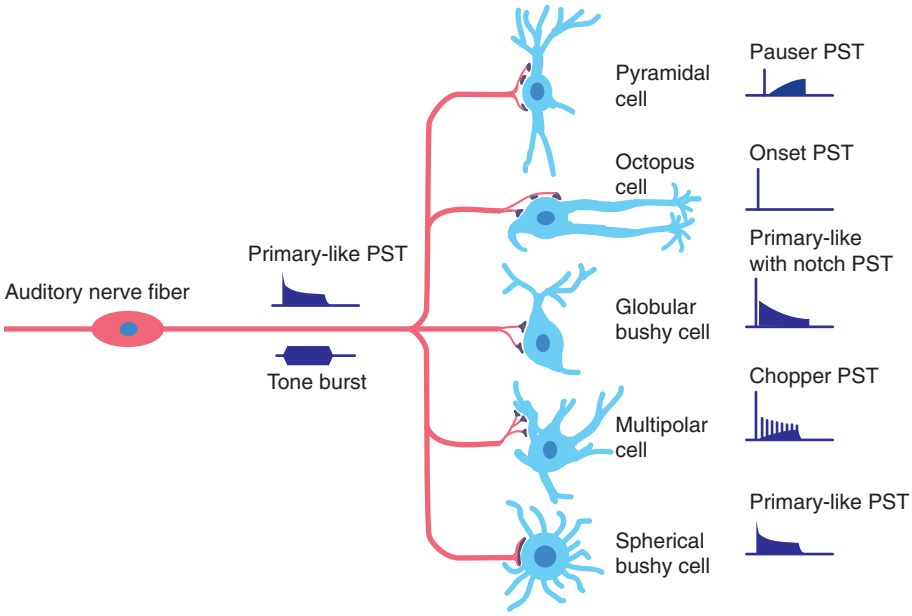


FIGURE 25.17 Schematic of the main anatomical cell types of the cochlear nucleus and their corresponding poststimulus time (PST) histograms. (Left) An auditory nerve fiber is shown with its typical response, a primary-like PST histogram (shown in Fig. 25.15). (Center) The auditory nerve fiber divides to innervate the main cochlear nucleus cell types. (Right) PST histograms corresponding to these cell types are shown. In their PST histograms, pauser units fire an initial spike and then have a distinct pause before a slow resumption of activity. Onset units fire mainly at the tone burst onset. Primary-like units get their name from the similarity of their PSTs to those of primary auditory nerve fibers, but the primary-like with notch type additionally has a brief notch following the initial peak. Chopper units have regular interspike intervals that result in regular peaks in the PST. Most of these patterns are very different from the primary-like PST and irregular interspike intervals of the auditory nerve fibers. For histograms, the sound stimulus is typically a 25-ms tone burst with frequency at the CF of the neuron and sound level at 30 dB above threshold. *From Kiang (1975).*

such as multipolar cells, are confined to a second region (nucleus angularis). The bushy cells in nucleus magnocellularis have excellent phase locking, but a poor dynamic range of response when sound level is increased. Multipolar neurons in nucleus angularis have complementary properties: poor phase locking, but excellent dynamic ranges. Overall, these results suggest that in the auditory system of diverse animal species there are separate neural pathways for time coding versus sound level coding as well as for other types of coding. Thus, the cochlear nucleus is the level where parallel pathways in the auditory system begin.

Classification by Response Map

Another important way to classify neurons in the cochlear nucleus, as well as throughout the auditory nervous system, is by the response map of a neuron (Fig. 25.19). Response maps are plotted on graphs of sound level versus frequency; like tuning curves, these maps show areas of excitation, but they additionally show areas of inhibition. They are especially valuable where inhibitory influences play a role in shaping responses, such as in the dorsal subdivision of the cochlear nucleus, the inferior colliculus, and at higher stages of the auditory system. In the dorsal cochlear nucleus, five response types have been defined (Fig. 25.19). Type I neurons have excitatory tuning curves similar to those of the auditory nerve (Fig. 25.12) and have no inhibitory areas. Other response types have progressively larger inhibitory areas; for example, type IV neurons have only a small excitatory area near CF and a narrow, knife-sharp excitatory area above CF, with the rest of their area dominated by inhibition. This inhibition is generated by inhibitory circuits within the dorsal cochlear nucleus, most likely from type II neurons. Type IV neurons correspond to the main projection neurons of the dorsal cochlear nucleus, the pyramidal neurons, whereas type II neurons are likely to be inhibitory interneurons. One hypothesis for the functional role of type IV neurons is that their sharp borders between excitatory and inhibitory areas serve to detect the spectral notches that result from the acoustic characteristics of the external ear, especially the pinna. These notches could be used for sound localization, as their frequencies depend on sound source location. The notches also depend on position of the pinna; in fact, type IV neurons receive input from brainstem somatosensory nuclei that may inform the type IV neurons about the position of the animal's moveable pinna. Animals that lack a moveable pinna, such as humans and cetaceans (whales and porpoises), have a dorsal cochlear nucleus that differs greatly from that of other mammals by being unlayered and possibly lacking in several cell types (granule and cartwheel cells).

BOX 25.3

GIANT SYNAPTIC TERMINALS: ENDBULBS AND CALYCES

The largest synaptic terminals in the brain are contained in the central auditory pathway. There are two types of these giant synaptic terminals: (1) endbulbs of Held, which are found in the ventral cochlear nucleus (Fig. 25.18A), and (2) calyceal endings, which are found in the medial nucleus of the trapezoid body. Calyces are so large that it is possible to use patch electrodes to record and clamp the presynaptic terminal while simultaneously doing the same with their postsynaptic target in *in vitro* preparations (Forsythe, 1994). This type of study has given insight into presynaptic and postsynaptic regulation of transmitter release at this glutamatergic synapse.

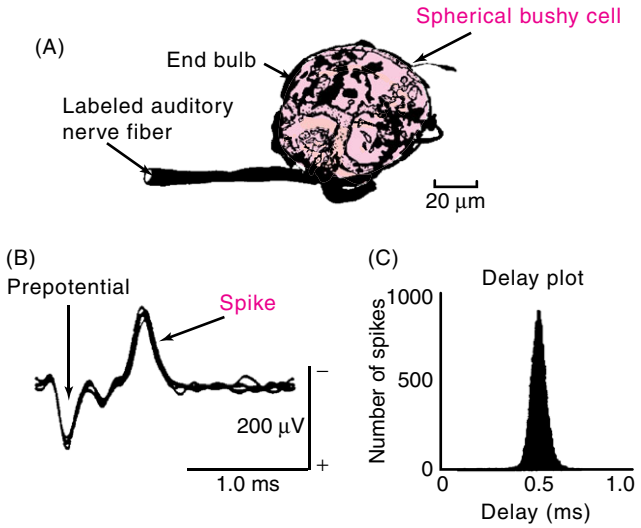


FIGURE 25.18 (A) Drawing of a labeled auditory nerve fiber forming an endbulb of Held that completely envelops a spherical bushy cell in the anteroventral cochlear nucleus. *From Rouiller et al. (1986).* (B) Superimposed waveforms recorded by an extracellular metal electrode in the anteroventral cochlear nucleus. The endbulb is such a large synaptic ending that it generates a prepotential, which is followed after a synaptic delay by the spike from the bushy cell. *From Pfeiffer (1966).* (C) Histogram of the delay time between prepotential and spike. This represents the synaptic delay between endbulb and bushy cell. The delay of this synapse has an exceptionally low jitter in time. *From Molnar and Pfeiffer (1968).*

Endbulbs and calyces enable secure transmission of information to their postsynaptic neurons. Endbulbs of Held are formed by primary auditory nerve fibers; each nerve fiber forms one or sometimes two endbulbs. The endbulbs contact and completely encircle their postsynaptic target, the spherical bushy cells of the cochlear nucleus (Fig. 25.18A). The endbulb probably forms hundreds of synapses directly onto the soma of the spherical bushy cell. In single-unit recordings near bushy cells, an extracellular microelectrode records a complex waveform that differs from recordings in other regions of the brain (Fig. 25.18B). The waveform consists of a prepotential followed about 0.5 ms later by a spike. The prepotential is likely from the presynaptic endbulb and the spike is from the postsynaptic bushy cell. The delay between the two events has a low jitter, or variability, in time (Fig. 25.18C).

The large influence of the endbulb would, for most neurons, last many milliseconds so that two closely spaced presynaptic spikes would produce only one postsynaptic spike or a second spike that was delayed, thus smearing in time the pattern of input spikes. However, the bushy cell membrane potential recovers quickly because its membrane contains specialized K^+ channels that allow the cell to repolarize after firing an impulse so it is "reset" and ready to fire again (Manis & Marx, 1991). The overall effect of endbulb–bushy cell specializations are to replicate the spike pattern of the auditory nerve fiber, producing a PST histogram and phase-locking pattern like that of an auditory nerve fiber. These characteristics are important because the end-bulb–bushy cell synapse is the only central synapse in the pathway to the medial superior olivary nucleus, a nucleus where timing information from the two ears is compared in order to localize sound sources (Fig. 25.21A).

M. Christian Brown

References

Forsythe, I. D. (1994). Direct patch readings from identified presynaptic terminals mediating glutamatergic EPSCs in the rat CNS, *in vitro*. *Journal of Physiology, 479,* 381–387.

Manis, P. B., & Marx, S. D. (1991). Outward currents in isolated ventral cochlear nucleus neurons. *Journal of Neuroscience, 11,* 2865–2880.

Classification by Laterality of Response: The Response to the Contralateral versus Ipsilateral Ear

A final important way to classify neurons is by their laterality of response, which is defined as whether the neuron responds to the contralateral or ipsilateral ear and whether the response is excitatory or inhibitory (reviewed by Irvine, 1986). Many neurons in central auditory nuclei above the cochlear nucleus are binaural and can be influenced by sound presented to either ear. A predominant pattern, however, is for the neuron to be excited by sound in the contralateral ear (the side opposite to where the neuron is located). The influence of the ipsilateral ear can be excitatory, inhibitory, or mixed. The contralateral

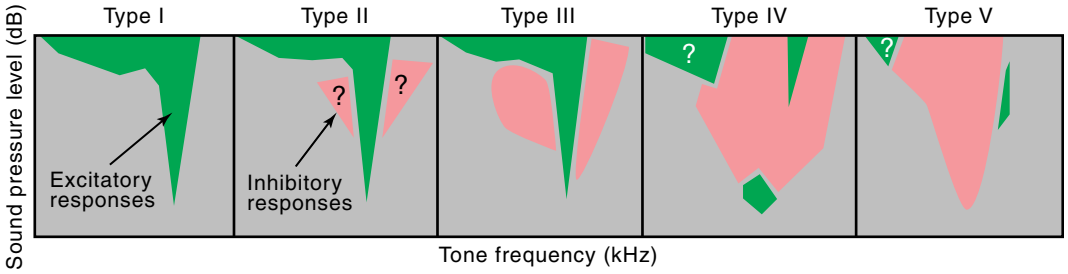


FIGURE 25.19 Unit classification by "response map." Shown are response areas for the five response types (types I–V) typically found in the cochlear nucleus. Response types are distinguished by the positions of their response areas: excitatory (green shading) and inhibitory (pink shading). Question marks show variable or uncertain areas. Type I neurons have excitatory response areas similar to the tuning curves of auditory nerve fibers. Type II neurons have similar excitatory areas and are inferred to have inhibitory flanking areas because they have little response to broadband signals like noise. Type III neurons have similar excitatory areas and definite inhibitory areas on either side. Type IV neurons have a small excitatory area at low levels (near the characteristic frequency), as well as a knife-sharp excitatory area at higher frequencies. Inhibition dominates much of the remainder of their response area. Type V neurons are similar to type IV neurons but lack a low-level excitatory area. *From Young (1984).*

response results from the fact that many central auditory pathways cross to the opposite side of the brain (Fig. 25.16). There are also uncrossed pathways; these pathways generate the response to the ipsilateral ear. Despite an influence of the ipsilateral ear, lesion studies indicate the functional importance of excitation from the contralateral ear. For instance, damage to the inferior colliculus or auditory cortex on one side of the brain decreases the ability to localize sound sources within the spatial hemifield on the opposite side of the body. Thus, as in other sensory and in motor systems, one side of the brain is primarily concerned with function on the opposite side of the body.

The Auditory Brainstem Uses Binaural Cues to Determine Sound Location

An important function of the auditory system is to determine the location of sound sources in space. Binaural sound localization occurs when cues from both ears are used to locate sounds (reviewed by Wightman & Kistler, 1993). This is the most accurate type of localization for determining the azimuthal position of a sound source. Binaural localization uses two cues: interaural time differences (ITDs) and interaural level differences (ILDs) (Fig. 25.20). ITDs result because sound reaches the ear closest to the source sooner than the ear farther from the source (Fig. 25.20A). ITDs are a good cue for localization because they depend greatly on the azimuth of the source. For ongoing sounds, such as pure tones, they can be translated into phase differences in the sound waveforms at the two ears. These phase differences are useful at low frequencies, but become ambiguous for high frequencies (above about 1.5 kHz in humans) because by the time sound reaches the ear away from the source, the waveform has repeated by a cycle or more. The decline in phase locking for frequencies above 1 kHz is a second reason that phase differences are less important for localizing sounds at high frequencies.

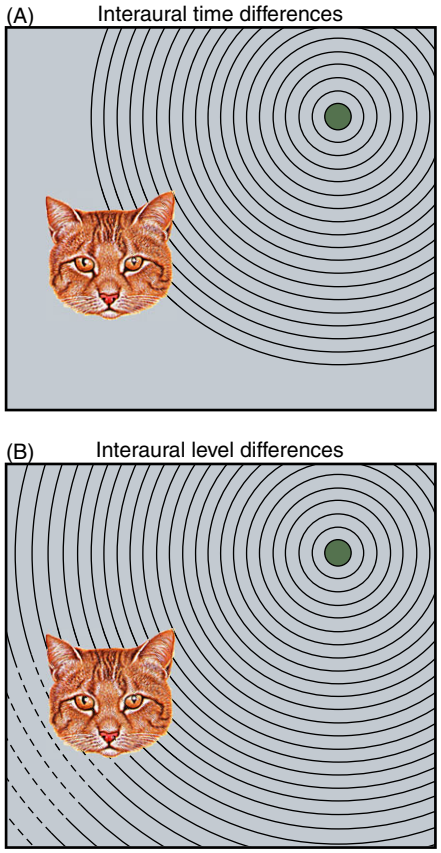


FIGURE 25.20 The two cues for binaural localization of sound. A sound source is shown as a solid dot to the right of the animal's head and sound waves are shown as concentric lines. (A) Interaural time differences result from the longer time it takes sound to travel from the source to the ear away from the source. (B) Interaural level differences result from the head forming a "sound shadow," reducing the level of sound at the ear away from the source.

Interaural level differences (ILDs) result when the head forms a "sound shadow," reducing the level of sound at the ear away from the source (Fig. 25.20B). ILDs vary greatly with sound-source azimuth, but due to the directional

characteristics of sound, ILDs are significant only at high frequencies and are much smaller at low frequencies. Thus, for sound localization at low frequencies (<1 kHz), ITDs are the major cues, but for localization at high frequencies (>3 kHz), ILDs are the major cues. For human performance using pure tones, the accuracy of azimuthal localization is good at low frequencies and at high frequencies, but somewhat less accurate at middle frequencies, perhaps because the cues are more ambiguous in this range. In psychophysical experiments in humans under optimal conditions, the minimum discriminable angle for localization of a sound source approaches one degree of azimuth. The physical cues corresponding to this angle are about 10 μs in ITD or 1 dB in ILD. When experiments are conducted with headphones to manipulate the ITDs and ILDs, such small interaural differences are indeed discriminable in human subjects.

Mechanisms of Interaural Time Sensitivity

The auditory pathway has more brainstem nuclei than other sensory pathways, and much of this processing is devoted to determining the location of a sound source. Two neural circuits that provide sensitivity to ITD or ILD are within the superior olivary complex (Fig. 25.21; Irvine, 1986; Yin & Chan, 1990). The neural mechanisms that generate sensitivity to ITDs are impressive given the small sizes of the differences. For instance, the central nervous system must be able to detect ITDs of 10 μs by comparing spikes coming in from the neural channels of the two sides that differ in time by 10 μs. However, 10 μs is less than the rise time of neural spikes, making the incoming spikes almost identical. The neural circuit that is sensitive to ITDs in the mammal is the medial superior olive (MSO) and its inputs (Fig. 25.21A); a similar circuit is present in birds in nucleus laminaris. The MSO inputs are from primary-like units (spherical bushy cells) of the left and right cochlear nuclei. These inputs preserve the phase-locking and timing characteristics of the auditory nerve because of the low jitter in the endbulb/bushy cell synapse (see Box 25.3). For a low frequency sound with an ITD, phase-locked spikes from one side will have a time difference relative to the other side. Phase-locked spikes will repeat this time difference many times during the waveforms of a continuous sound. Neural delay lines that "make up" for this time difference are formed by axons for both contralateral and ipsilateral inputs, in the model originally proposed by Lloyd Jeffress (Fig. 25.21A). An axon forms a delay line simply because it takes time for an impulse to travel along the axon. Within the MSO, neurons respond best when they receive coincident input from the two sides; thus a neuron in the middle of the drawing of the MSO in Figure 25.21A would respond best if the delays were about equal. A neuron at the bottom of the drawing has a long contralateral axonal delay that makes up for stimuli coming in later from the ipsilateral side. Thus, this neuron would respond best if the sound to the contralateral ear were leading, which occurs for a sound source located on the contralateral side (Fig. 25.21B). Medial superior olive neurons are thus tuned to a particular ITD and respond less to other ITDs. The Jeffress model has recently been challenged: new data in small mammals demonstrates that the best delays are outside the physiological range for the size of the animal's head and the neural inhibition in the MSO is not explained in the Jeffress model (Brand, Behrend, Marquardt, McAlpine, & Grothe, 2002). As mentioned earlier, ITDs are most important for sound localization at low frequencies and the MSO has a tonotopic organization that is composed predominantly of neurons with low characteristic frequencies.

Mechanisms of Interaural Level Sensitivity

A second neural circuit in the brainstem, within the lateral superior olive (LSO), generates responses that are sensitive to ILDs (Fig. 25.21C). Input to the LSO from the ipsilateral side is excitatory by way of spherical bushy cells of the cochlear nucleus. Input from the contralateral side is inhibitory. This input originates from globular bushy cells of the cochlear nucleus and synapses by way of giant calyceal endings on neurons in the medial nucleus of the trapezoid body (MNTB; Fig. 25.21C). Most of the neurons in the medial nucleus of the trapezoid body are inhibitory and use the neurotransmitter glycine, which is a common inhibitory transmitter throughout the auditory brainstem. These inhibitory neurons then project to the LSO. Because of these inputs, LSO neurons compare the difference in levels of the sound at the two ears: they are excited when sound in the ipsilateral ear is of higher level, but are inhibited when sound in the contralateral ear is of higher level (Fig. 25.21D). When the sound is of equal level in the two ears, the strong contralateral inhibition usually dominates and there is little response. These neurons are thus excited by sound sources located on the ipsilateral side of the head. As the lateral superior olive projects further centrally, this ipsilateral-side response is transformed to a contralateral-side response by crossing to the inferior colliculus on the opposite side. (There is also an uncrossed projection, but it is inhibitory.) As mentioned earlier, ILDs are most important for sound localization at high frequencies and the LSO has a tonotopic organization that is composed predominantly of neurons with high characteristic frequencies.

Inferior Colliculus

Ascending input from lower brainstem centers converges at the inferior colliculus, which is a synaptic station for almost all ascending neurons. The inferior colliculus consists of several subdivisions, the best studied of which is the large, tonotopically organized central nucleus. In the central nucleus, direct input from the cochlear nucleus interacts with binaurally responsive input from the MSO

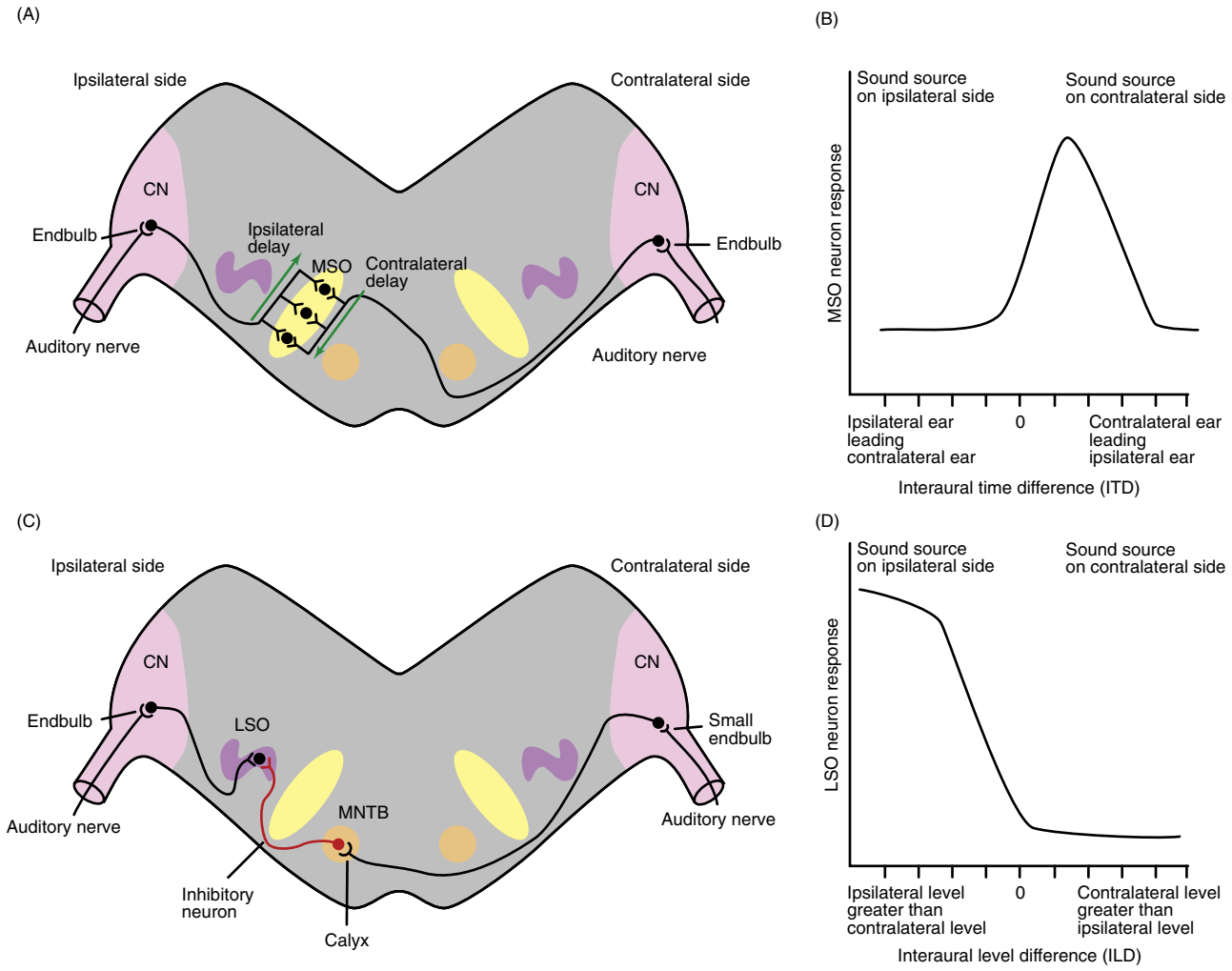


FIGURE 25.21 Innervation schematics and responses of two circuits in the lower brainstem that are important in binaural sound localization. Neuronal cell bodies are shown as dots, and fiber pathways are shown as lines; positions of large synaptic terminals (endbulbs and calyces) are indicated. (A) Circuit of the medial superior olive (MSO), which is sensitive to interaural time differences (ITD). Input to the cochlear nucleus (CN) from the auditory nerve terminates at the large endbulbs of Held that synapse onto spherical bushy cells (see Fig. 25.18A). Bushy cells project bilaterally such that a single MSO receives input from both sides. Bushy cell inputs form delay lines such that ITD is mapped along the MSO. Data suggest that the delay line is oriented rostrocaudally and that only contralateral inputs are delayed. (B) Response of an MSO neuron as a function of ITD. Neurons within the MSO respond when spikes from their two inputs arrive at the same time. The response plotted is of a neuron in the lower part of the MSO drawn in part A; there is a large response when the ipsilateral input lags so that early contralateral input has time to proceed down the axonal delay line to reach the neuron at the same time as the lagging ipsilateral input. This type of lagging ipsilateral input would be produced by a sound source located in the hemifield of space contralateral to the MSO where the recording is made. (C) Circuit of the lateral superior olive (LSO), which is sensitive to interaural level differences (ILD). Excitatory input arises from the ipsilateral CN. Inhibitory input (red line) from the contralateral side is through the medial nucleus of the trapezoid body (MNTB), a nucleus of inhibitory neurons. The large synaptic ending in the MNTB is called a calyx (similar to the endbulb pictured in Fig. 25.18). (D) Response of an LSO neuron as a function of ILD. There is a large response when sound is of higher level on the ipsilateral side and no response when sound is of higher level on the contralateral side. Thus, a response is produced by a sound source located in the ipsilateral hemifield.

and LSO. Low CF input (including that from the MSO) projects to the dorsolateral part of the colliculus and high CF input (including that from the LSO) projects to the ventromedial part. Many low CF collicular neurons are thus sensitive to ITDs like their MSO inputs, whereas many high CF collicular neurons are sensitive to ILDs like their LSO inputs. Interestingly, large lesions of the superior olivary complex do not completely disrupt ILD sensitivity in the colliculus; this is evidence that ILD sensitivity is created anew at levels above the LSO. ILD sensitivity may be created in part by the dorsal nucleus of the lateral lemniscus, a nucleus within the lateral lemniscus that sends a large inhibitory projection to the colliculus. ILD sensitivity may also be created anew by inhibitory mechanisms within the colliculus. Additional circuits for the generation of ITD sensitivity have not been identified,

thus the colliculus appears to be sensitive to ITD because of its inputs from the MSO.

An important question is whether the colliculus uses its inputs to form a "space map"—a mapping of sound source location to position within the brain. Such a mapping has not been reported in the mammalian inferior colliculus, but a mapping exists in the deep layers of the superior colliculus, where it is in register with a mapping of the visual field. These deep layers of the superior colliculus have sensorimotor functions concerning orientation movements of the head, eyes, and pinnae. Another auditory space map has been observed in the barn owl, which hunts for prey in darkness using acoustic cues (reviewed by Cohen & Knudsen, 1999). In the owl, the mapping has been observed within a nucleus homologous to the external nucleus of the inferior colliculus (nucleus mesencephalicus lateralis dorsalis). Here, spatial receptive fields of neurons are narrow in both azimuth and elevation and are arranged such that there is a mapping of sound source location within the brain. This mapping demonstrates that space maps do exist, but they do not appear to be a common feature of the auditory system of nonspecialized mammals.

The Medial Geniculate and Auditory Cortex Are the Highest Stages of the Auditory Pathway

The medial geniculate and auditory cortex contain a number of subdivisions that have clear tonotopic organization (reviewed by Clarey, Barone, & Imig, 1992; de Ribaupierre, 1997). For instance, in the cortex of the cat, there are four fields that have clear tonotopic mappings (fields AI, A, P, VP) (Fig. 25.22). The tonotopic axis of field AI runs from high frequencies rostrally to low frequencies caudally; other adjacent fields (such as A) have mirror-image tonotopy (Fig. 25.22B). The major ascending pathway that connects tonotopic areas at the highest levels of the pathway begins in the central nucleus of the inferior colliculus, forms synapses in the laminated ventral division of the geniculate and then continues to most of the tonotopic cortical fields. Other parallel pathways exist that connect regions with less tonotopic organization. One of these pathways begins in the dorsal cortex of the colliculus, forms synapses in the dorsal division of the medial geniculate and projects mainly to cortical field AII. Another, a polysensory pathway, begins in the external and dorsal nuclei of the colliculus and projects via the medial division of the geniculate to almost all the auditory cortical fields.

In general, the physiology of tonotopic areas has been explored most extensively. Units from the tonotopic areas, such as the ventral division of the geniculate and cortical field AI, tend to have short latencies and sharply tuned tuning curves. Neurons in areas with less obvious tonotopy tend to have longer latencies, broader tuning curves, and

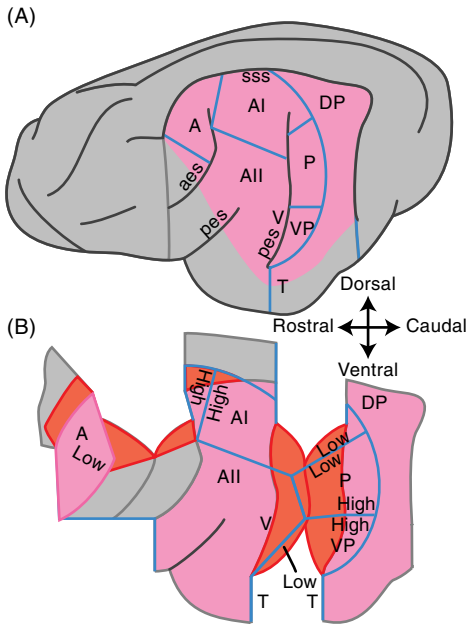


FIGURE 25.22 Auditory cortical fields in the temporal cortex of the cat. (A) Lateral view. (B) Lateral view that is "unfolded" to show the parts of the fields that are normally hidden within the sulci (orange shading), as well as the high- and low-frequency limits of the tonotopic fields. The four tonotopic fields are the anterior (A), primary (AI), posterior (P), and ventroposterior (VP). Positions of the lowest and highest CFs in these fields are indicated in B. Note that at the boundaries of the tonotopic fields, the direction of tonotopy is reversed so that adjacent fields have "mirror-image" tonotopy. Other cortical fields have less rigidly organized or little tonotopy. These fields are secondary (All), ventral (V), temporal (T), and dorsoposterior (DP). Also indicated are suprasylvian sulcus (sss) and anterior and posterior ectosylvian sulci (aes, pes). *From Imig and Reale (1980).*

responses that can habituate or stop responding after multiple presentations of the stimulus, especially in anesthetized preparations. In humans, the primary auditory cortex is located on Heschl's gyrus, which is on the superior surface of the temporal lobe. Sound-evoked activation is seen in this region with positron emission tomography (PET) and functional magnetic resonance imaging (fMRI).

The medial geniculate exerts its influence on ascending information, but it does so with extensive influence from the cortex. The geniculate receives a large number of projections from the auditory cortex: it probably receives more input from cortex than from lower centers. One large projection is from the tonotopic cortical fields AI and A to the tonotopic ventral division of the medial geniculate. Apparently, however, the geniculate mediates important functions that do not require the cortex. For example, fear conditioning is a behavior that can be established by pairing a sound with a painful electric shock in rats. After conditioning is established, conditioned responses such as an increase in blood pressure can be elicited by sound alone. Lesions of the auditory pathway up to and including the geniculate have a large effect on such conditioning;

however, lesions of the auditory cortex do not produce large alterations (LeDoux, Sakaguichi, & Reis, 1984). Pathways directly from the geniculate to the amygdala mediate such conditioned responses. Cortical neuron responses are also changed by conditioning; for example, the pairing of an acoustic stimulus with a noxious stimulus greatly alters the response of cortical neurons. Such observations indicate that responses at these high levels of the auditory pathway are context dependent.

Cortical Fields and Sound Localization

Physiological and behavioral studies suggest a role for cortical field AI in sound localization. In AI, many neurons are sensitive to the spatial position of a sound source. As expected from the fact that auditory pathways are predominantly crossed, a large number of cortical neurons respond to sound sources centered on the contralateral side. Behaviorally, sound localization performance is decreased by lesions of AI, the posterior auditory field (a tonotopic field just caudal to the primary auditory field), and an area near the anterior ectosylvian sulcus (Malhotra & Lomber, 2006). Furthermore, the deficit is frequency specific: if the lesion is in areas of AI tuned to certain frequencies, the deficit is observed only for those frequencies (Jenkins & Merzenich, 1984).

Coding of Complex Signals in the Cortex

Animals emit a variety of vocal signals that make possible a wide range of behaviors. Studies in echolocating bats demonstrate specialized cortical fields and responses (see Box 25.4). Cortical neurons show selectivity to complex sounds such as species-specific calls over what would be predicted from their pure tone frequency selectivity (Rauschecker, Tian, & Hauser, 1995). For instance, a neuron that responds well to a particular type of call does not usually respond as well to the call played backward in time even though it has an identical frequency content. Similar results in response to speech stimuli have been

BOX 25.4

BAT ECHOLOCATION

Bats offer unique opportunities for researchers studying the auditory system (rev. by Popper and Fay, 1995). Echolocating bats emit high-frequency pulses of sound and use the return echoes to locate and capture their flying insect prey. These pulses are above the upper frequency limit of human hearing. They were discovered in the 1940s by Donald Griffin, who was then an undergraduate student at Harvard University. The echolocation performance of the bat is impressive: bats are adept enough to find a mosquito above a golf course at night. Indeed, Griffin demonstrated that bats can avoid wires as thin as 0.3 mm diameter while flying in a darkened room.

There are more than 800 species of echolocating bats, all within the suborder Microchiroptera. One of the best-studied bats is the mustached bat, *Pteronotus parnellii*. It emits echolocating pulses that have an initial part of constant frequency, followed by a part of decreasing frequency (frequency modulated). This bat is thus called a CF-FM bat (Fig. 25.23B). Some other species of bats emit only the FM portion and are called FM bats. Bats use differences between the emitted pulse and the returned echo to locate targets. Relative to the pulse, the returned echo is delayed because of the sound's round-trip travel time from the bat to the target. The delay between the pulse and echo can thus be used as a measure of target range. The echo is changed in frequency, or Doppler-shifted, if the bat is moving relative to the reflecting surface. Bats that are moving toward a reflecting surface will have an echo that is Doppler-shifted toward higher frequency. Most background surfaces will be Doppler shifted about the same. However, targets that are moving differently from the background will generate different Doppler shifts. For instance, a moth with moving wings will reflect an echo with a moving Doppler shift that is quite different from the stationary background. Such differences are presumably used by the CF-FM bat to locate moving targets. Insects are not always passive in the face of such predatory behavior. For instance, some moths have good ultrasonic hearing and take evasive flight maneuvers when exposed to a bat's pulse.

Research on the auditory system of bats has been greatly aided by knowledge of the "relevant" stimulus, since much of the time the bat hears an emitted pulse followed a short time later by an echo. Pioneering work by Nobuo Suga has demonstrated that the auditory cortex of the mustached bat contains many specialized areas (Fig. 25.23C). One large area (DSCF area) is devoted to processing the Doppler-shifted echo for the strongest harmonic of the pulse, near 60 kHz. Within the DSCF area, there is a mapping of neuronal best level for a response as well as the best or characteristic frequency. In other cortical regions, there are neurons whose response properties cause them to detect various features of the pulse and echo. For instance, neurons in the FM-FM region respond preferentially to two FM pulses separated by a specific delay. The best delays for these neurons range from 0.4 to 18 ms, which would correspond to target ranges of 7 to 310 cm. Furthermore, these best delays are mapped along

BOX 25.4 (*cont'd*)

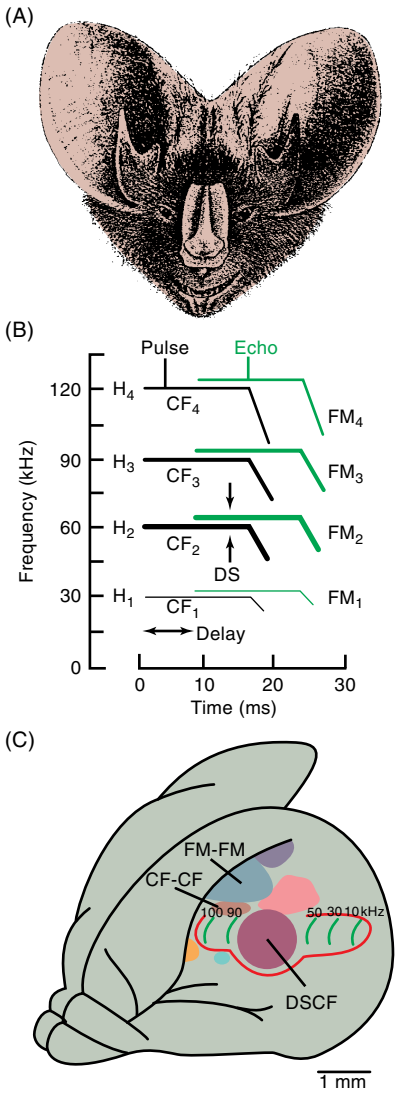


FIGURE 25.23 (A) Close-up view of an echolocating bat, *Megaderma lyra*. Like many echolocating bats, this bat has an enlarged nose leaf that probably focuses the echolocating pulse ahead. Also like many echolocating bats, this bat has enlarged external ears that make the bat more sensitive to echoes coming from ahead. *From Griffin (1958).* (B) Schematic of emitted pulse and returned echo of the mustached bat, *Pteronotus parnellii*. The emitted pulse consists of four harmonics (H_1–H_4), the strongest of which is H_2 at about 60 kHz. Each harmonic has an initial part of constant frequency (CF) and a later part of changing frequency (frequency modulation, FM). The echoes are returned after a travel time that causes a delay relative to the pulse. Additionally, if the target is moving relative to the bat, the echo is returned with a Doppler shift (DS) in frequency. (C) Dorsolateral view of the auditory cortex of the mustached bat showing several of the areas specialized for processing the echolocation signals. The primary auditory cortex, AI, is delineated by a red line and its isofrequency contours are indicated in kHz. It contains a region called the Doppler-shifted CF region (DSCF), which is a greatly expanded region devoted to the most prominent component of the Doppler-shifted echo (60 to 63 kHz). Other shaded regions contain neurons that are combination sensitive and respond to combinations of the pulse and echo, often at certain delays. Neurons that respond to pulse/echo CF portions are in the CF/CF region. Those that respond to pulse/echo FM portions are in the FM/FM region. *From Fitzpatrick et al. (1993).*

cortical distance. Work on the bat cortex has resulted in specific hypotheses about the function of most of their auditory cortical fields. Our state of knowledge of the cortical fields in other animals is by contrast much more primitive.

M. Christian Brown

Reference

Popper, A. N., & Fay, R. R. (Eds.), (1995). *Hearing by bats. Springer handbook of auditory research.* New York, NY: Springer-Verlag.

found in recordings of units from auditory areas in the superior temporal gyrus of human patients undergoing surgery for epilepsy.

Human speech is an especially complex acoustical signal. Language is processed in perisylvian cortex including Wernicke's area, which is close to auditory cortex and can be considered an auditory association area (see Chapter 49). Imaging studies demonstrate the processing of such complex acoustic stimuli in areas near primary auditory cortical fields. Further studies will be very useful in showing more specifically the distinct functions of the auditory fields and how they work together to process sound stimuli.

Summary

Much about the central auditory pathways remains to be explored. Although these pathways are mapped tonotopically, whether there are other mappings in orthogonal dimensions is unknown. Much current research is focused on the plasticity of central auditory responses with auditory experience and after hearing loss and whether this plasticity is generally present at the many subcortical auditory centers. The fact that there are descending systems at many levels of the pathways is known, but research is necessary to show how these systems alter the processing of auditory information.

References

Brand, A., Behrend, O., Marquardt, T., McAlpine, D., & Grothe, B. (2002). Precise inhibition is essential for microsecond interaural time difference coding. *Nature, 417,* 543–547.

Brownell, W. E., Bader, C. R., Bertrand, D., & de Ribaupierre, Y. (1985). Evoked mechanical response of isolated hair cells. *Science, 227,* 194–196.

Clarey, J. C., Barone, P., & Imig, T. J. (1992). Physiology of thalamus and cortex. In A. N. Popper & R. R. Fay (Eds.), *The mammalian auditory pathway: neurophysiology* (pp. 232–334). New York, NY: Springer-Verlag.

Cohen, Y. E., & Knudsen, E. I. (1999). Maps versus clusters: Different representations of auditory space in the midbrain and forebrain. *Trends in Neurosciences, 12,* 128–135.

Corey, D. P., & Hudspeth, A. J. (1983). Kinetics of the receptor current in bullfrog saccular hair cells. *Neuroscience, 3,* 962–976.

Corwin, J. T., & Cotanche, D. A. (1988). Regeneration of sensory hair cells after acoustic trauma. *Science, 240,* 1772–1774.

Dallos, P., Wu, X., Cheatham, M. A., Gao, J., Zheng, J., Anderson, C. T., Zuo, J., et al. (2008). Prestin-based outer hair cell motility is necessary for mammalian cochlear amplification. *Neuron, 58,* 333–339.

de Ribaupierre, F. (1997). Acoustical information processing in the auditory thalamus and cerebral cortex. In G. Ehret & R. Romand (Eds.), *The central auditory system* (pp. 317–397). New York, NY: Oxford Univ. Press.

Ehret, G., & Romand, R. (Eds.), (1997). *The central auditory system*. New York, NY: Oxford Univ. Press.

Evans, E. F. (1975). Cochlear nerve and cochlear nucleus. In W. D. Keidel & W. D. Neff (Eds.), *Handbook of sensory physiology* (vol. 5, pt. 2, pp. 1–108). Berlin, Germany: Springer Verlag.

Fitzpatrick, D. C., Kanwai, J. S., Butman, J. A. & Suga, N. (1993). Combination-sensitive neurons in the primary auditory cortex of the mustached bat. *Journal of Neuroscience, 13*, 931–940.

Flock, Â. (1965). Transducing mechanisms in the lateral line canal organ receptors. *Cold Spring Harbor Symposia on Quantitative Biology, 30*, 133–145.

Geisler, C. D. (1998). *From sound to synapse*. Oxford, NY: Oxford University Press.

Griffin, D. R. (1958). *Listening in the dark*. New Haven, CT: Yale University Press.

Hudspeth, A. J., & Corey, D. P. (1977). Sensitivity, polarity, and conductance change in the response of vertebrate hair cells to controlled mechanical stimuli. *Proceedings of National Academy of Science of the United States of America, 74,* 2407–2411.

Imig, T. J. & Reale, R. A. (1980). Patterns of cortico-cortical connections related to tonotopic maps in cat auditory cortex. *Journal of Comparative Neurology*, 192, 293–332.

Irvine, D. R. F. (1986). *The auditory brainstem*. Berlin: Springer-Verlag.

Kiang, N. Y. S. (1975). Stimulus representation in the discharge patterns of auditory neurons. In D. B. Tower (Ed.), *The nervous system, vol. 3: Human communication and its disorders* (pp. 81–96). New York, NY: Raven Press.

Kiang, N. Y. S. (1984). Peripheral neural processing of auditory information. In J. M. Brookhart & V. B. Mountcastle (Eds.), *Handbook of physiology* (vol. 3, pp. 639–674). Bethesda, MD: American Physiological Society.

Kiang, N. Y. S. & Peake, W. T. (1988). Physics and physiology of hearing. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey & R. C. Luce (Eds.), *Stevens' handbook of experimental psychology, vol. I: Perception and motivation* (pp. 277–326), Hoboken, NJ: John Wiley & Sons.

LeDoux, J. E., Sakaguchi, A. & Reis, D. J. (1984). Subcortical efferent projections of the medial geniculate nucleus mediate emotional responses conditioned to acoustic stimuli. *Journal of Neuroscience, 4*, 683–698.

Liberman, M. C. (1982). The cochlear frequency map for the cat, Labeling auditory-nerve fibers of known characteristic frequency. *Journal of the Acoustical Society of America, 72,* 1441–1449.

Lindsay, P. H., & Norman, D. A. (1972). *Human information processing: An introduction to psychology*. New York, NY: Academic Press.

Loeb, G. E. (1985). The functional replacement of the ear. *Scientific American, 252*, 104–111.

Lonsbury-Martin, B. L., & Martin, G. K. (1990). The clinical utility of distortion-product emissions. *Ear and Hearing, 11,* 144–154.

Malhotra, S., & Lomber, S. G. (2006). Sound localization during homotopic and heterotopic bilateral cooling deactivation of primary and non-primary auditory cortical areas in the cat. *Journal of Neurophysiology, 97,* 26–43.

Merchant, S. N., & Nadol, J. B., Jr. (Eds.), (2010). *Schuknecht's pathology of the ear*. Shelton, CT: People's Medical Publishing House-USA.

Molnar, C. E. & Pfeiffer, R. R. (1968). Interpretation of spontaneous spike discharge patterns of neurons in the cochlear nucleus. *Proceedings of the IEEE, 56*, 993–1004.

Parks, T. N., Rubel, E. W., & Fay, R. R. (Eds.), (2004). *Plasticity of the auditory system. Springer handbook of auditory research*. New York, NY: Springer Science+Business Media, LLC.

Pfeiffer, R. R. (1966). Anteroventral cochlear nucleus: Wave forms of extracellularly recorded spike potentials. *Science, 154*, 667–668.

Pickles, J. O., Comis, S. D., & Osborne, M. P. (1984). Cross-links between stereocilia in the guinea-pig organ of Corti, and their possible relation to sensory transduction. *Hearing Research, 15,* 103–112.

Pickles, J. O. & Corey, D. P. (1992). Mechanoelectrical transduction by hair cells. *Trends in Neuroscience*, 15, 254–259.

Popper, A. N., & Fay, R. R. (Eds.), (1992). *The mammalian auditory pathway: Neurophysiology*. New York, NY: Springer-Verlag.

Rauschecker, J. P., Tian, B., & Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. *Science, 268,* 111–114.

Rhode, W. S., & Greenberg, S. (1992). Physiology of the cochlear nuclei. In A. N. Popper & R. R. Fay (Eds.), *The mammalian auditory pathway: Neurophysiology* (pp. 94–152). New York, NY: Springer-Verlag.

Robertson, D., & Irvine, D. R. F. (1989). Plasticity of frequency organization in auditory cortex of guinea pigs with partial unilateral deafness. *Journal of Comparative Neurology, 282,* 456–471.

Rose, J. E., Hind, J. & Anderson, D. J. (1971). Some effects of stimulus intensity on response of auditory nerve fibers in the squirrel monkey. *Journal of Neurophysiology, 34,* 685–699.

Rouiller, E. M., Cronin-Schreiber, R., Fekete, D. M. & Ryugo, D. K. (1986). The central projections of intracellularly labeled auditory nerve fibers in cats: An analysis of terminal morphology. *Journal of Comparative Neurology, 249*, 261–278.

Ruggero, M. A. (1992). Physiology and coding of sound in the auditory nerve. In A. N. Popper & R. R. Fay (Eds.), *The mammalian auditory pathway: Neurophysiology* (pp. 34–93). New York, NY: Springer-Verlag.

Russell, I. J. & Sellick, P. M. (1983). Low-frequency characteristics of intracellularly recorded receptor potentials in guinea-pig cochlear hair cells. *Journal of Physiology, 338*, 179–206.

Rybalchenko, V., & Santos-Sacchi, J. (2003). Cl- flux through a non-selective, stretch sensitive conductance influences the outer hair cell motor of the guinea pig. *Journal of Physiology, 547,* 873–891.

Ryugo, D. K., Popper, A. N., & Fay, R. R. (Eds.), (2011). *Auditory and vestibular efferents*. New York, NY: Springer Science+Business Media, LLC.

Santos-Sacchi, J. (1991) Reversible inhibition of voltage-dependent outer hair cell motility and capacitance. *Journal of Neuroscience, 11*, 3096–3110.

Santos-Sacchi, J. (1992). On the frequency limit and phase of outer hair cell motility: Effects of the membrane filter. *Journal of Neuroscience, 12,* 1906–1916.

Santos-Sacchi, J., Song, L., Zheng, J., & Nuttall, A. L. (2006). Control of mammalian cochlear amplification by chloride anions. *Journal of Neuroscience, 26,* 3992–3998.

Warr, W. B., Guinan, J. J. Jr. & White, J. S. (1986). Organization of the efferent fibers: The lateral and medial olivocochlear systems.

In R. A. Altschuler, D. W. Hoffman & R. P. Bobbin (Eds.), *Neurobiology of hearing: The cochlea* (pp. 333–348). New York, NY: Raven Press.
Wightman, F. L., & Kistler, D. J. (1993). Sound localization. In W. A. Yost, A. N. Popper & R. R. Fay (Eds.), *Human psychophysics* (pp. 155–192). New York, NY: Springer-Verlag.
Winslow, R. & Sachs, M. B. (1987). Effect of electrical stimulation of the crossed olivocochlear bundle on auditory nerve response to tones in noise. *Journal of Neurophysiology, 57,* 1002–1021.
Yin, T. C. T., & Chan, J. C. K. (1990). Interaural time sensitivity in medial superior olive of cat. *Journal of Neurophysiology, 64,* 465–488.
Young, E. D. (1984). Response characteristics of neurons of the cochlear nuclei. In C. I. Berlin (Ed.), *Hearing science: Recent advances* (pp. 423–460). San Diego: College-Hill Press.
Zheng, J., Shen, W., He, D. Z., Long, K. B., Madison, L. D., & Dallos, P. (2000). Prestin is the motor protein of cochlear outer hair cells. *Nature, 405,* 149–155.

Suggested Readings

Dallos, P., Popper, A. N., & Fay, R. R. (Eds.), (1996). *The cochlea*. New York, NY: Springer-Verlag.
Jahn, A. F., & Santos-Sacchi, J. (Eds.), (2001). *Physiology of the ear* (2nd ed.). New York, NY: Raven Press.
Pickles, J. O. (2008). *An introduction to the physiology of hearing* (3rd ed.). London: Academic Press.
Popper, A. N., & Fay, R. R. (Eds.), (1995). *Hearing by bats*. New York, NY: Springer-Verlag.
Von Bekesy, G. (1960). *Experiments in hearing*. New York, NY: McGraw-Hill.

M. Christian Brown and Joseph Santos-Sacchi

CHAPTER

26

Vision

OVERVIEW

Vision is the most studied and perhaps the best understood topic in sensory neuroscience. This chapter is concerned primarily with vision in mammals and focuses on the pathway in the visual system that has to do with perception: from the retina to the lateral geniculate nucleus of the thalamus (LGN) and on to the multiple areas of visual cortex. Other regions of the brain that receive visual input (such as the superior colliculus) are dealt with briefly in this chapter and more extensively in the chapter on eye movements (see Chapter 32).

Studying vision provides the opportunity to explore the brain at many different levels, from the physical and biochemical mechanisms of phototransduction to the boundary between psychology and physiology (Chapter 44). At each of these levels, the visual system has evolved to solve a number of difficult problems. In terms of the physical stimulus, vision operates over extremely wide ranges of illumination. The visual system detects single photons in the dark but can also see clearly in bright sunlight, when the retina is bombarded with over 10^{14} photons per second. At a much higher level of complexity, ensembles of neurons in the cerebral cortex are able to solve extremely difficult problems, such as extracting the three-dimensional motion of an object from two-dimensional retinal images.

At every moment, the visual system is confronted with the vast amount of information present in visual scenes. The complex circuitry of the retina has evolved so that much of this information is extensively processed and relayed to the rest of the central nervous system, both efficiently and with great fidelity. Vision, however, has not evolved to treat all of this information equally; instead, it appears to be best suited to extract the sort of information that may be useful to animals, including humans, in a natural environment. Vision allows animals to navigate in the world; to judge the speed and distance of objects; and to identify food, members of other species, and familiar or unfamiliar members of the same species.

In many animals, primates in particular, more of the brain is devoted to vision than to any other sensory function. This is perhaps because of the extreme complexity of the task required of vision: to classify and to interpret the wide range of visual stimuli in the physical world. At the highest levels of processing, the cerebral cortex extracts from the world the diverse qualities experienced as visual perception: from motion, color, texture, and depth to the grouping of objects, defined by the combination of simple features.

The Receptive Field Is the Fundamental Concept in Visual Physiology

The strategies that the brain uses to solve the problems of vision can be understood at a very intuitive level. The most useful concept to aid this intuition is that of the *receptive field*, which is the cornerstone of visual physiology. As defined by H. K. Hartline in 1938 (Ratliff, 1974, p. 167), a visual receptive field is the "region of the retina which must be illuminated in order to obtain a response in any given fiber." In this case, "fiber" refers to the axon of a retinal neuron, but any visual neuron, from a photoreceptor to a visual cortical neuron, has a receptive field. The definition was later extended to include not only the region of the retina that excited a neuron, but also the specific properties of the stimulus that evoked the strongest response. Visual neurons can respond preferentially to the turning on or turning off of a light stimulus—termed *on-and-off* responses—or to more complex features, such as color or the direction of motion. Any of these preferences can be expressed as attributes of the receptive field.

© 2013 Elsevier Inc. All rights reserved.

SENSORY SYSTEMS DETECT CONTRAST OR CHANGE

In the 1930s and 1940s, Hartline developed the concept of the receptive field with studies of the axons of individual neurons that project from the lateral eye of the horseshoe crab *(Limulus)* and from the frog's eye (Ratliff, 1974). The lateral eye of the *Limulus* is a compound eye made up of about 300 *ommatidia* arranged in a roughly hexagonal array. Each ommatidium contains optical elements, photoreceptors, and a single neuron whose axon joins the optic nerve. Hartline found that when an isolated ommatidium was illuminated, the firing rate of its axon increased. More surprisingly, the firing of the same axon was decreased by a light stimulus in any adjacent ommatidium. This form of antagonistic behavior, known as *lateral inhibition,* serves to enhance responses to edges while reducing responses to constant surfaces. Without it, visual neurons would be just as sensitive to a featureless stimulus, such as a clean white wall, as to stimuli defined by edges, such as a white square on a black wall. Similar spatially antagonistic visual responses were found in mammals, as first demonstrated by Kuffler (1953) in the retina of the cat (Box 26.1; Fig. 26.3).

Lateral inhibition represents the classic example of a general principle: most neurons in sensory systems are best adapted for detecting changes in the external environment. This principle can be explained in behavioral terms. As a rule, it is change that has the greatest significance for an animal—for example, the edge of an extended object or a static object beginning to move. This principle can also be explained in terms of information processing. Given a world that is filled with constants—with uniform objects, with objects that move only rarely—it is most efficient to respond only to changes.

Several types of visual responses can be discussed in terms of the detection of change, or of *contrast:* defined as the fractional difference in luminance between two

BOX 26.1

KUFFLER'S STUDY OF CENTER-SURROUND RETINAL GANGLION CELLS

The classic experiments that Kuffler (1953) performed on retinal ganglion cells have formed the foundations for much of the subsequent physiological analysis of the mammalian visual system. Even beyond the study of vision, they represent a model for understanding the neurobiology of sensory systems. These experiments were performed *in vivo* in anesthetized cats. The first step in this sort of experiment is the careful placement of a fine microelectrode close to a single neuron so that action potentials can be recorded extracellularly. An oscilloscope trace of the firing pattern of this neuron is important, but the sound of action potentials on an audio monitor is even more critical. This immediate feedback allows the researcher to search for visual stimuli that excite or inhibit the neuron. Kuffler's first finding was that there were two categories of ganglion cells, as Hartline had seen in the retina of the frog. The cells were either *on,* i.e., excited by light increment, or *off*—that is, excited by light decrement.

One of Kuffler's most important contributions was the careful mapping of lateral interactions in the retina, or what he termed the *center-surround* structure of the receptive field. When an *on* ganglion cell was being studied, a small light spot placed in the center of its receptive field would cause an immediate increase in firing rate of the cell (Fig. 26.3B, top). The center of the receptive field of an *on* ganglion cell was defined as all positions where the small spot evoked an *on* excitatory response. When the same spot was placed just beyond the center, in the region termed the *surround*, the neuron decreased its firing rate (Fig. 26.3B, bottom). Kuffler mapped the spatial extent of the regions that evoked excitation or inhibition simply by listening to the responses to spots flashed at many different locations. Alternatively, Kuffler studied receptive fields by searching for an optimal stimulus—one that increased the firing of a ganglion cell most effectively. The strongest stimulus for an *on* center cell was a spot of light that filled the receptive field center entirely (Fig. 26.3C). Similarly, the most effective inhibitory stimulus was a bright annulus shown to the surround alone. Following such strong inhibition, the cell had an excitatory response when the stimulus was turned off (Fig. 26.3D). Finally, Kuffler studied interactions between receptive field subregions. A large bright stimulus that covered both center and surround was found to evoke a much weaker response than a smaller spot confined to the center; the surround inhibition weakened or altogether eliminated the central excitation (Fig. 26.3E).

Kuffler's early experiments, along with those of Hartline, established the technical and conceptual foundations for the field of visual physiology. Almost all subsequent work in this field can be seen as falling into the three broad categories of experiments, exemplified by Kuffler's 1953 study: (1) the mapping of responses with isolated, suboptimal stimuli, (2) the search for an optimal stimulus, and (3) the study of interactions between responses evoked by two or more stimuli.

R. Clay Reid

stimuli. There are several forms of contrast. The first is spatial contrast, the detection of which is enhanced by neurons in the retina that have lateral inhibition (center-surround organization, see Box 26.1). Next, there is temporal contrast, or change over time. Starting in the retina, visual neurons are affected very little by slow changes in illumination but are extremely sensitive to more rapid changes. Finally, there is motion, which is distinguished by characteristic changes in a stimulus over both space and time. Many neurons in the visual system are excited selectively by objects that have a certain rate or direction of motion. In summary, contrast sensitivity can take on at least three forms: sensitivity to spatial variations in a stimulus (spatial contrast), sensitivity to changes over time (temporal contrast), and sensitivity to changes in both space and time (motion).

Receptive Fields Encode Increasingly High-Order Features of the Visual World

From the photoreceptors to the multiple visual cortical areas, the visual system is hierarchical. One level provides input to the next in a feedforward progression, although lateral interactions and feedback are almost always present as well. As a general rule, receptive fields at successive stages of processing (from photoreceptors, bipolar cells, ganglion cells, and geniculate neurons, through neurons in multiple visual cortical areas) encode increasingly high-level features of the visual stimulus. The outer segment of a photoreceptor, which contains the visual pigment, is influenced only by a small point in visual space. It is therefore almost entirely insensitive to the spatial structure of a stimulus. At the opposite extreme, neurons in the inferior parietal region of visual cortex seem to respond best when the animal is viewing a specific face (see Chapter 50).

These two extremes of visual responses illustrate the visual system's dual task: to maintain generality—the ability to respond to any stimulus—while also being able to represent specific, environmentally important classes of stimuli. High-level neurons classify visual stimuli by integrating information that is present in the earlier stages of processing, but also by ignoring information that is independent of that classification. For instance, motion-sensitive neurons in area MT of the cerebral cortex (see later) are exquisitely sensitive to the direction and rate of motion of an object, but very poor at distinguishing the object's color or its position. This lack of localization is quite common in high-level neurons: receptive fields become larger as the features they represent become increasingly complex. Thus, for instance, neurons that respond to faces typically have receptive fields that cover most of visual space. For these cells, large receptive fields have a distinct advantage: the preferred stimulus can be identified no matter where it is located on the retina.

Summary

The mammalian visual system is a complex, hierarchical system that can be studied from a number of different viewpoints. This chapter concentrates on the physiology of visual neurons, which is based on the study of receptive fields. Originally, receptive fields were defined as the area of the retina that could evoke responses in a visual neuron. The concept has evolved to include the stimulus attributes that lead to a neural response, such as color, motion, or even the complex features of a specific physical object. Within the hierarchy of the visual system—from retina to the multiple areas of the visual cortex—neurons are selective for increasingly complex or high-order features.

THE EYE AND THE RETINA

The Optics of the Eye Project an Inverted Visual Image on the Retina

The study of vision begins with the eye (Fig. 26.1), whose refractive properties are determined by the curvature of the cornea and the lens behind it. These optical elements act to focus an inverted image on the retina, where the first stages of neural visual processing take place. The curvature of the cornea is fixed, but the curvature of the lens is adjusted by smooth muscles that flatten

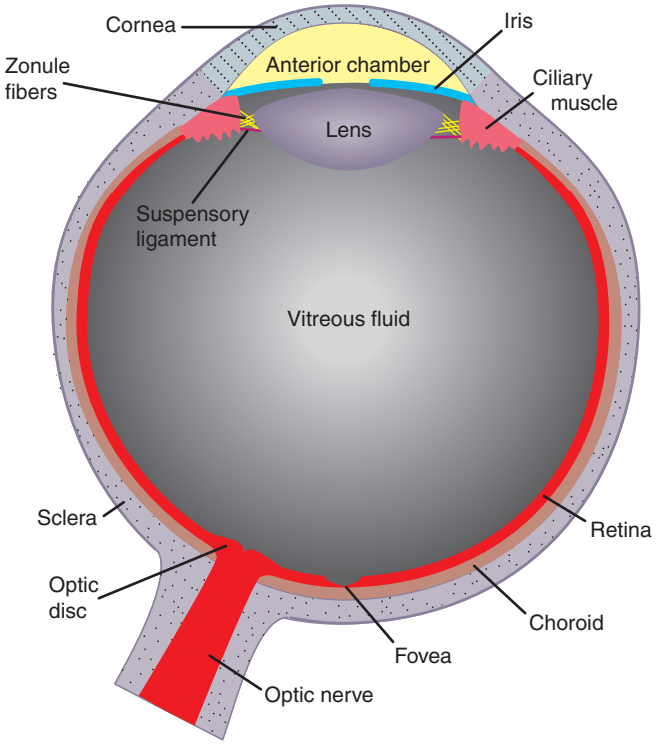


FIGURE 26.1 Schematic diagram of the human eye.

the lens when they relax, thus bringing more distant objects into focus. The amount of light that reaches the retina is controlled by the iris, whose aperture is the pupil. The iris, which is situated between the cornea and the lens in the *anterior chamber* of the eye, contracts at high light levels, and expands in the dark.

The Retina Is a Three-Layered Structure with Five Types of Neurons

The anatomy of the retina has an almost crystalline beauty, a beauty that is enhanced by the clear relationships between form and function (Dowling, 1997). It is composed of five principal layers: three layers of cell bodies separated by two layers of neural processes, dendrites and axons (Fig. 26.2). The vertebrate retina is oriented within the eye so that light must travel through the entire thickness of the neuropil to reach the photoreceptors.

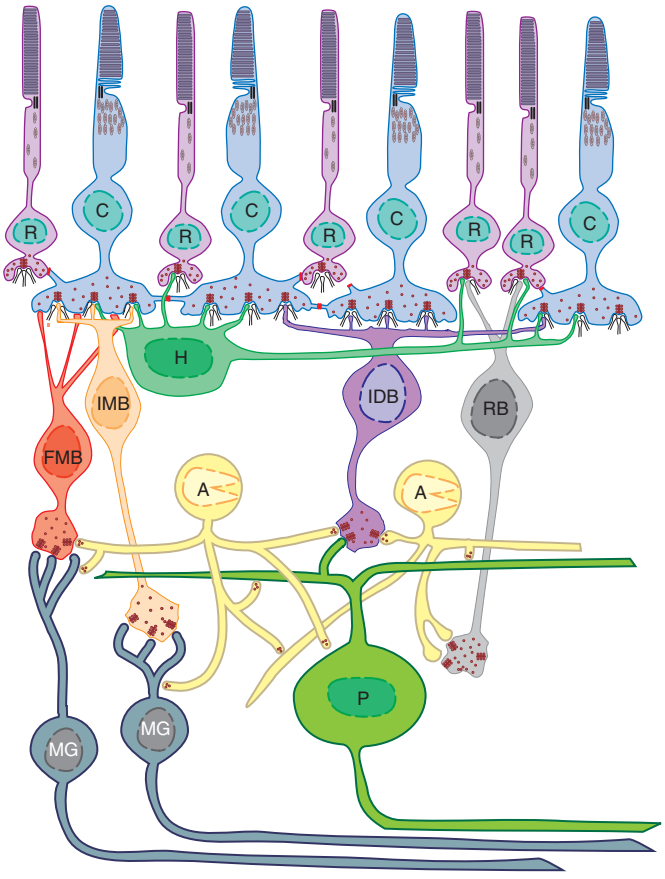


FIGURE 26.2 Summary diagram of the cell types and connections in the primate retina. R, rod; C, cone; H, horizontal cell; FMB, flat midget bipolar; IMB, invaginating midget bipolar; IDB, invaginating diffuse bipolar; RB, rod bipolar; A, amacrine cell; P, parasol cell (also confusingly called an M cell because of its thalamic targets, see text for details); MG, midget ganglion cell (also confusingly called a P cell). *Adapted from Dowling (1997).*

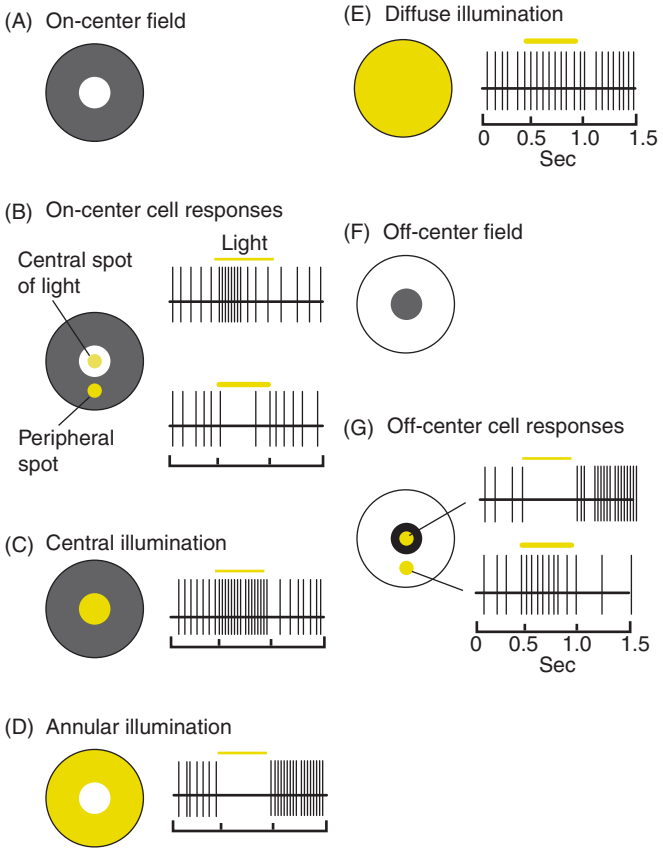


FIGURE 26.3 Visual responses of *on*-center (white) and *off*-center (dark gray) retinal ganglion cells. Visual stimuli are indicated in yellow, and the responses to these stimuli are shown to the right. See text for details.

Of the three cell layers, the first is farthest from the center of the eye and is thus called the *outer nuclear layer*. It contains the cell bodies of the photoreceptors, the rods, and the cones. The next cell layer is the *inner nuclear layer*, which contains the cell bodies of the interneurons of the retina, both excitatory and inhibitory. These include horizontal cells, bipolar cells, and amacrine cells. Finally, the *ganglion cell layer* is home to the retinal neurons whose axons form the optic nerve, the sole pathway from the retina to the rest of the central nervous system. Interposed between the cell body layers are two layers of cell processes: *inner* and *outer plexiform layers*. The two plexiform layers are the sites of all interactions between the neurons of the retina.

The retina is one of the few circuits in the nervous system simple enough that cell types and connections can be learned without great effort. This is made even easier, since the role of each of the five cell types can be placed within a simple functional scheme. The two main attributes of the output of the retina—the point-to-point representation of the visual image and the spatially antagonistic center-surround interactions in the receptive field

(see Box 26.1)—can be understood in terms of the anatomy. The direct pathway, photoreceptor → bipolar cell → ganglion cell, is the substrate for the center of the receptive field of the ganglion cell and thus for its spatial resolution. Lateral interactions in the retina, most notably the center-surround antagonism, or lateral inhibition, are mediated by horizontal cells and amacrine cells.

Photoreceptors Are Hyperpolarized by Light

The two major types of photoreceptors in the vertebrate eye are the rods and cones (Fig. 26.2). Both types of photoreceptor have an outer segment that contains the molecular machinery for phototransduction, an inner segment that contains densely packed mitochondria, a cell body that contains the nucleus and other important organelles, and a terminal process that releases neurotransmitter. Although rods and cones were once viewed as the only cells in the vertebrate eye to transduce light energy into an electrical response, recent work shows that phototransduction also takes place in melanopsin-containing ganglion cells that likely convey nonspatial information about light levels to regions of the brain involved with circadian rhythms (Berson, 2003).

Rods and cones work together to allow the visual system to operate over a wide range of luminance conditions. Under scotopic conditions when luminance levels are very low (e.g., starlight), only the rods are active. As luminance levels increase to mesopic conditions (e.g., moonlight), both the rods and cones contribute to vision. As luminance levels increase further yet, to photopic conditions (e.g., sunlight), rod responses saturate and only the cones contribute to vision. Because the center of our retina (the fovea) contains very few if any rods, we have a foveal blind spot under very dim conditions. Consequently, a dim star will disappear when we move our eyes to look at it directly.

Phototransduction occurs when photons of light are absorbed by photopigments located in the outer segments of photoreceptors. Photopigments have an organic, light absorbing component (11-*cis* retinal) and a protein component (opsin). The opsin determines the wavelength specificity of the photopigment and has the ability to interact with G proteins. Rods contain a single photopigment, rhodopsin, while cones contain one of three cone opsins (the relationship between these opsins and color vision is discussed below). Once a photon is absorbed, a cascade of events occurs that ultimately affects the membrane potential of the photoreceptor. This cascade begins with the activation of the G protein (transducin) that then activates a phosphodiesterase that hydrolyzes cGMP and reduces the intracellular concentration of cGMP. Because the outer membrane of a photoreceptor contains many cGMP-gated cation channels, a decrease in the intracellular concentration of cGMP will cause the photoreceptor to hyperpolarize. In this way, light increments lead to hyperpolarization and a reduction in neurotransmitter release, while light decrements lead to depolarization and an increase in transmitter release. Importantly, photoreceptors do not produce action potentials, but rather have graded potentials that are modulated around a mean level.

Bipolar Cells Can Be Hyperpolarized or Depolarized by Light

Similar to photoreceptors, bipolar cells generate graded potentials rather than action potentials. Bipolar cells are divided into two broad classes, termed *on* and *off* bipolars, that respond to light stimuli with depolarization and hyperpolarization, respectively. *On* bipolars are also known as *invaginating bipolars*, and *off* bipolars are known as *flat bipolars* because of the shape of the synaptic contacts they receive from photoreceptors (Fig. 26.2). Because all photoreceptors use glutamate as their neurotransmitter, these opposite responses require that *on* and *off* bipolar cells respond to the same neurotransmitter in opposite ways (Wu, 1994). Glutamate, typically an excitatory transmitter, is inhibitory for *on* bipolars. It acts by closing a cGMP-gated sodium channel, similar to that seen in photoreceptors. The depolarization of *on* bipolars by light results not from excitation, but from the removal of inhibition, which occurs when photoreceptors are hyperpolarized by light. *Off* bipolars are excited by glutamate via more typical kainate receptors (named for the pharmacological agent that selectively activates them; see Chapter 9). When these cells are inhibited by light, the inhibition is in fact due to the removal of tonic excitation, which again occurs when the photoreceptors are hyperpolarized.

The functional importance of the *on* and the *off* pathways can best be understood in terms of contrast. From the bipolar cells onward—that is, once *on* and *off* pathways have been established—visual neurons respond best to spatial and temporal contrast rather than to absolute light levels. Objects in the world are visible by the light they reflect; their borders are discerned usually because of different degrees of reflectance, which leads to contrast (changes in illumination, except for shadows, tend to be much more gradual). Because the visual system is adapted for seeing objects that can be either brighter or darker than their backgrounds, it is not too surprising that it is equally sensitive to positive and negative contrast steps.

Most ganglion cells, the output cells of the retina, receive their main excitatory input from bipolar cells. Thus, these ganglion cells also have either *on* or *off* responses in the center of their receptive fields, according to which class of bipolar cells provide their input (see Box 26.1). Other ganglion cells, known as *on-off* cells, respond to both light onset and to light offset. These latter cells receive input from both classes of bipolar cells. Many of these cells respond preferentially to stimuli moving in a particular direction of motion.

Horizontal and Amacrine Cells Mediate Lateral Interactions in the Retina

There is a complex three-way synaptic relationship among the terminals of the photoreceptors, the processes of horizontal cells, and the dendrites of bipolar cells (Fig. 26.2). Horizontal cells are inhibitory (GABAergic), and, as the name implies, they contact photoreceptors over a much larger horizontal extent of retina than bipolar cells. Thus, because they are the only neurons that sample over a sufficiently large area in the outer plexiform layer, horizontal cells are thought to be responsible for the antagonistic surround seen in the receptive fields of bipolar cells.

The second family of lateral interneurons in the retina, the amacrine cells (literally, cells with no axons), is a diverse class of cells that exhibits a wide range of morphologies. As is true of horizontal cells, all processing takes place on the dendrites of amacrine cells (in the inner plexiform layer), which contain both pre- and postsynaptic elements (Fig. 26.2). Most amacrine cells use GABA, glycine, or both acetylcholine and GABA as neurotransmitters, but it is thought that virtually every transmitter found in the brain is used by some amacrine cell. The functions of most of these diverse subclasses remain poorly understood. In addition to a possible role in the creation of the antagonistic surround, it has been proposed that some amacrine cells are responsible for the nonlinear responses in Y cells of the cat (see Box 26.2; Fig. 26.3) or for the direction selectivity of certain classes of ganglion cells identified in the rabbit and rodent retina.

BOX 26.2

QUANTITATIVE METHODS IN THE STUDY OF VISUAL NEURONS CLASSIFICATION OF RETINAL GANGLION CELLS

Although neuroscientists have learned a tremendous amount about the visual system using the tools developed by Kuffler, researchers have been following a parallel line of studies of the visual system by asking different sorts of questions with the aid of more quantitative methods. In Kuffler's experiments and, most notably, many of Hubel and Wiesel's (see later), easily produced stimuli (such as spots, edges, and bars) were used to stimulate visual neurons. Action potentials were detected primarily with an oscilloscope and audio monitor. In the more quantitative studies of visual physiology, stimuli are shown primarily on video monitors and data are recorded with computers. The quantitative study of visual physiology constitutes a large field, which stems from the work of Hartline, Ratliff, Campbell, Barlow, and others (see Ratliff, 1974). Analysis of the receptive fields of retinal ganglion cells in the cat performed by Enroth-Cugell and Robson (1966) represents an early and influential example of this line of research.

Stimuli used by Enroth-Cugell and Robson, and in many studies that followed theirs, consisted of *gratings*: light and dark bands whose luminance varied in a sinusoidal fashion across a screen. Gratings were not designed to be the most effective stimuli for visual neurons. The broad goal of this sort of experiment is to study not just what makes neurons respond best, but to study how they would respond to *any* stimulus and to probe what mechanisms they use to produce these responses. The analytical framework that goes along with these experiments is called *systems theory*, the study of input-output (or stimulus-response) systems. *Linear systems*, those that simply add up all of their inputs to produce an output, constitute an important part of this theory.

Enroth-Cugell and Robson used a simple test of linearity in their study of retinal ganglion cells in the cat. A grating was presented at different positions (called phases), and the responses to its introduction and removal were recorded. Two questions were addressed for each cell studied: (1) If a certain phase (whose bars were arranged white, black, and white; as in Fig. 26.4, 0°) excited the cell, did the opposite grating (black, white, and black; Fig. 26.4, 180°) inhibit it? (2) More importantly, were there null positions for the grating that evoked no response, as excitation and inhibition were perfectly balanced (Fig. 26.4, 90 and 270°)?

One class of ganglion cells, called X *cells*, passed these tests for linear spatial summation (Fig. 26.4, left). A second class of cells, *Y cells*, behaved differently. Like X cells, these cells also had a center-surround receptive field when mapped with spots and annuli. When studied with two opposite gratings, however, the responses evoked were not equal and opposite. Instead, the introduction and removal of the gratings at all positions resulted in two peaks of excitation, and no null position could be found (Fig. 26.4, right).

The use of systems theory and quantitative techniques in the study of the visual system has had a number of notable successes. It has helped in the classification of different families of visual neurons, such as X and Y cells. Further, it has been used to elucidate some of the mechanisms responsible for the responses of visual neurons (such as directionally selective cells). Perhaps most importantly, along with a large body of psychophysical experiments that employ the same stimuli, it has helped create a common framework in which the responses of neurons can be related to perception.

R. Clay Reid

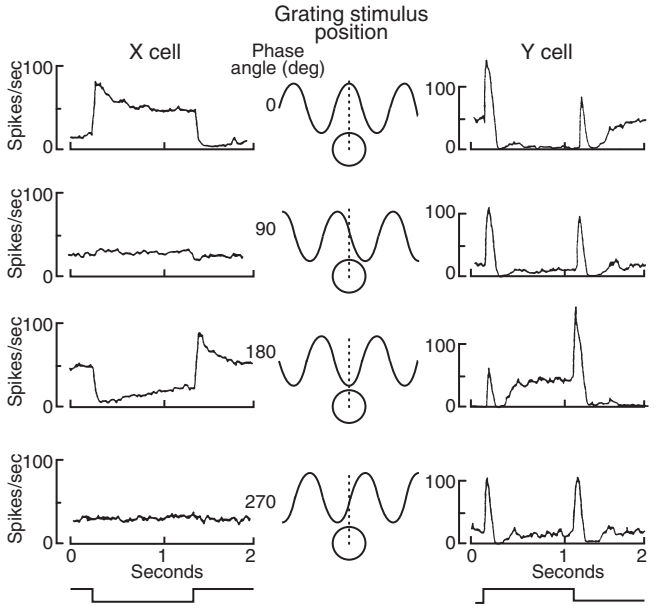


FIGURE 26.4 Visual responses of X and Y cells to contrast-reversing sine-wave gratings at different spatial phases (different positions, indicated schematically in the middle column). Visual responses, in spikes per second, are shown as poststimulus time histograms synchronized to the repetitive stimulus. Upward deflection of the lowest trace indicates introduction of the grating pattern (contrast on), and downward deflection indicates removal of the pattern (contrast off, but no change in mean luminance). *Adapted from Enroth-Cugell and Robson (1966).*

Retinal Ganglion Cells Provide the Output of the Retina

Visual information leaves the eye via the optic nerve, which is composed of the axons of all of the different classes of ganglion cells. The optic nerve begins at the *optic disc* (Fig. 26.1). Because there are no photoreceptors at the optic disc, this circular region constitutes a blind spot in the retina. The routing of ganglion cell axons to different parts of the brain provides a number of interesting problems for developmental neurobiology (Chapters 15, 16, and 21). The routing of axons is both macroscopic—different types of ganglion cells project to different regions of the brain—and microscopic: within any given target region, axons are sorted out in a precise manner. In each of two principal targets of retinal axons, the superior colliculus and lateral geniculate nucleus of the thalamus (LGN), a topographical map of visual space is created in which spatial relations are maintained between neighboring neurons.

This chapter is concerned primarily with the classes of retinal ganglion cells that project to LGN neurons, which in turn project to the primary visual (striate) cortex. Although cells in the LGN are more than mere relays of information from the retina, their receptive fields are quite similar to those of their ganglion cell inputs. The following discussion of retinal receptive fields, therefore, can serve equally well to describe receptive fields in the LGN.

Parallel Pathways Are Composed of Distinct Classes of Retinal Ganglion Cells in the Cat

Kuffler found that retinal ganglion cells in the cat have a stereotyped center-surround organization. Similar cells (both *on* and *off*) were found at every position across the entire retina. Researchers later discovered, however, that there are in fact several different functional classes of ganglion cells, each with distinct response properties. Although different cell classes have been found in different species, in none is the retina composed of a single mosaic of identical cells. The fact that there are multiple types of output from the retina has had a profound effect on the study of visual processing in the brain. The degree to which these *parallel pathways* (Livingstone & Hubel, 1984; Merigan & Maunsell, 1994; Sincich & Horton, 2005) are combined or kept separate has been a major theme in the study of central visual pathways, particularly the visual cortex.

The idea that there are parallel sensory pathways into the brain is not new. At the beginning of the nineteenth century, both Bell and Müller argued for what Müller termed the "specific energy of nerves." This term is a precursor of the idea that axon sensory neurons are "labeled lines," each of which conveys distinct signals to the brain. A surprisingly modern version of this idea was given in 1860 by Helmholtz in the *Handbook of Physiological Optics* (2000), in which he expressed Thomas Young's theory of color vision (1807) in the following manner: "The eye is provided with three distinct sets of nervous fibers. Stimulation of the first excites the sensation of red, stimulation of the second the sensation of green, and stimulation of the third the sensation of violet."

In one of the earlier studies using a quantitative approach to receptive field mapping (see Box 26.2), Enroth-Cugell and Robson (1966) found that cat retinal ganglion cells could be classified into two categories: X and Y (more categories have been found subsequently). This classification was based on a simple criterion: did ganglion cells simply add together all of their excitatory and inhibitory inputs (linear summation), or were the interactions between inputs more complicated (nonlinear summation)? The linear/nonlinear distinction between X and Y cells (see Box 26.2; Fig. 26.4) may seem fairly abstract, but these cells turned out to be different in a number of other ways. Most notably, Y cell receptive fields are, on average, three times larger than those of neighboring X cells. X cells are therefore far more numerous, because many more X cell receptive fields are needed to effectively tile, or cover, the retina. One view is that the large, nonlinear receptive fields of Y cells make them well suited to

detect change, at the cost of their ability to signal the exact location or nature of the stimulus; similarly, X cells are better at localizing more specific stimulus features, but less sensitive in detecting change.

Primate Ganglion Cells Project to Three Subdivisions of the Lateral Geniculate Nucleus: Parvocellular, Magnocellular, and Koniocellular

Primates have a number of different classes of retinal ganglion cells, each with distinct cellular morphology, projection patterns, and visual response properties. This discussion focuses on ganglion cells that project to the lateral geniculate nucleus (LGN). There are many differences among primate species (nocturnal versus diurnal, Old World versus New World); this section concentrates on the macaque—a diurnal, Old World primate. The macaque visual system has been studied extensively with anatomical, physiological, and behavioral methods. Most importantly, macaque vision is very similar to human vision.

The most numerous class of ganglion cells in the macaque retina are sometimes referred to as P *cells*, so called because they project to the dorsal-most layers of the LGN, the *parvocellular layers* (small cells; Fig. 26.7). M cells, which project to the two ventral layers of the LGN, the *magnocellular layers* (large cells), constitute a second class of retinal neurons. P and M cells have distinct morphologies (Fig. 26.2). P cells (also known as *beta* cells or, in the central retina, midget ganglion cells) have very small dendritic fields. M cells (also known as *alpha* cells or parasol cells) have much larger dendritic fields. Finally, members of a third class of retinal cells project to the *intercalated layers*, located between the parvocellular and magnocellular layers. These intercalated layers are also termed *koniocellular* (dust-like, or tiny cells). There are roughly 1,000,000 P cells in the retina and parvocellular neurons in the LGN; there are 100,000 M and magnocellular neurons. Although the intercalated layers appear quite sparse and thin in standard histological sections (Fig. 26.7), there are as many intercalated cells as there are magnocellular cells. Due to their size and location, intercalated cells have been difficult to study and little is known about them in the macaque. Therefore the following section discusses only the functional properties of P and M pathways.

P and M Pathways Have Different Response Properties

As noted earlier, the receptive field properties of the LGN relay cells match closely those of their retinal afferents, but the two parallel pathways, P → parvocellular and M → magnocellular, are quite different. Five main characteristics distinguish the responses of P cells and M cells. (1) P cell receptive fields are smaller than M cell receptive fields at the same retinal position. (2) M cell axons conduct impulses faster than P cell axons. (3) The responses of P cells to a prolonged visual stimulus, particularly a color stimulus, can be very sustained, whereas M cells tend to respond more transiently. (4) Most P cells

BOX 26.3

INHERITED AND ACQUIRED DEFECTS OF COLOR VISION: RETINAL AND CORTICAL MECHANISMS

Inherited defects of color vision have been studied for over 200 years, originally with psychophysical methods but more recently with the tools of molecular genetics. Normal color vision is *trichromatic* (literally, three colored) because there are three different classes of cone photoreceptors. In the most common form of color vision defect, there are three pigments, but one of them is abnormal, or anomalous. More severe defects are found in *dichromats*, who lack a single cone type entirely. The current terminology for the three types of dichromacy was proposed by von Kries in 1897: *protanopia* (literally, the first type of blindness, or "red-blindness"), *deuteranopia* (second type, or "green-blindness"), and *tritanopia* (third type, or "blue-blindness"). While dichromats are unable to make certain color distinctions, only the very rare *monochromats*—people with one cone type or only rods—are truly color blind.

The genetics of red-green defects have long been known to be X linked: they are inherited from the mother and are fairly common in men, but they are uncommon in women.

Because of the overlap in cone pigments, inherited color vision defects are not as simple as terms such as "red blindness" might suggest. In broad terms, protanopes and deuteranopes are unable to make specific discriminations along the red-green axis. Depending on the defect, they are unable to distinguish certain reds and greens from gray or from each other. Tritanopes are unable to discriminate between colors with and without a short wavelength component, such as between gray and certain shades of violet or yellow.

R. Clay Reid

BOX 26.4

VISION IN MICE

The choice of species is always of great importance in attacking any biological problem. This chapter has concentrated on the macaque monkey because its visual system is very similar to that of humans. But vision has been an important topic in neuroscience for the past 50 years, not merely because we wish to understand how humans see but because it is also a tractable system for addressing many fundamental topics, from development through neural coding. Although traditionally neglected in visual neuroscience, the mouse is quickly becoming one of the most studied species (Chalupa & Williams, 2008), as it is in much of neuroscience.

The rodent visual system has many homologies with the primate visual system (and most other mammals), but there are several clear differences. Most obviously, while mice are certainly not proverbially blind, they have far lower acuity than primates. Humans can discern features as small as 1 minute of arc, but mice are barely able to make discriminations at the scale of 1 degree, 60 times larger. The past decade has told us that there is a great richness of visual processing in the mouse retina, thalamus, and visual cortex, and we can learn a great deal from studying mouse vision. For instance, although we have long known that the visual world is analyzed in parallel by numerous different classes of ganglion cells, it is likely that we will soon have a definitive list of ganglion cells in the mouse. With mouse lines that target single classes of cells, there has been a spate of recent discoveries about ganglion cells, such as a previously neglected class, identified with a transgenic marking method, that respond preferentially to upward motion (Kim et al., 2008). The visual cortex of rodents has many similarities with other species—with orientation-selective neurons, such as simple cells. However, it also has one fundamental difference with most other species in that it has no map for stimulus orientation, a characteristic found also in rats and other rodents (Ohki et al., 2005; Fig. 26.14A). There are also at least nine higher-order visual areas, which have the added benefit that they are all on the cortical surface (Wang & Burkhalter, 2007; Fig. 26.14B), so they can all be imaged with optical techniques. Although there will be no exact homologies between these areas and the extrastriate regions found in primates, they offer a rich opportunity to study classical problems, such as cortical arealization and hierarchical processing, with all of the modern tools of molecular genetics and *in vivo* imaging.

R. Clay Reid

are sensitive to the color of a stimulus; M cells are not. (5) M cells are much more sensitive than P cells to low-contrast, black-and-white stimuli. These last two differences, which have strong implications for the functional role of these pathways, are discussed next.

Color-Selective Responses in the P Pathway Are Derived from Antagonistic Inputs from L and M Cones

Color vision depends on the distinct sensitivities of the three classes of cone photoreceptors to different wavelengths of light (see Box 26.3; the rods are active only under low light levels and are not involved in color vision). Although fine distinctions can be made between different wavelengths, it is important to emphasize that each of the light-sensitive pigments in the three cone classes is sensitive to a broad range of wavelengths. The spectral sensitivities of these pigments overlap considerably. Consequently, the names sometimes used for these pigments—red, green, and blue sensitive—are inaccurate. For instance, the "red" pigment is most sensitive to light whose wavelength is 564 nm, or yellow, although it is more sensitive than the other pigments to red light. A more accurate terminology, one that identifies the relative sensitivities of the three cone absorption spectra, is usually employed: long, middle, and short wavelength sensitive (or L, M, and S).

Because green light will excite both long and middle wavelength cones, the color green can be distinguished only by the fact that it excites middle wavelength cones more strongly than it excites long wavelength cones. This distinction can be made by neurons that are sensitive to the difference between the signals from two cone classes, or neurons that are *color opponent*. In the retina and LGN, there are two categories of color-opponent cells: red-green and blue-yellow (yellow is made of the sum of long and middle wavelength cone signals; Dacey, 2000). The most numerous color-opponent neurons in the primate retina are red-green opponent P cells. These cells receive antagonistic input from long and middle wavelength-sensitive (L and M) cones.

Near the fovea, the center of the retina, a P cell receives input from only one bipolar cell. This *midget bipolar* in turn receives input from a single cone (Fig. 26.2). This arrangement ensures not only that the center of the ganglion cell's receptive field is as small as possible, but also that the center receives input from only one cone type: L or M. Classically, the antagonistic surround of these P cells (or their counterparts in the LGN) has been thought to be dominated by the other main cone class, M or L, respectively (Wiesel &

Hubel, 1966), although the nature of the surround in red-green P cells has been somewhat controversial (Dacey, 2000).

M Cells Are Highly Sensitive to Contrast

Because of their marked color sensitivity and small receptive fields, P cells have long been thought to be involved with the ability to make color discriminations and to see the finest details. When tested with black-and-white stimuli, however, P cells responded to low contrast very poorly compared to M cells. In a study that used the quantitative methodology similar to Enroth-Cugell and Robson's (see Box 26.2), Kaplan and Shapley (1986) measured the responses of P and M cells to sine-wave gratings of various contrasts (Fig. 26.5). Perceptually, stimuli can be detected that have less than 1% contrast (i.e., 1% luminance deviation from the mean). Kaplan and Shapley (1986) found that M cells often responded well to contrasts of less than 5% and that they gave their strongest responses at contrasts as low as 20%. P cells tended to respond poorly to contrasts below 10%, a stimulus that is quite salient, and rarely reached their strongest responses below 64%. From this, Kaplan and Shapley concluded that the ability to see low contrasts is due primarily to the M cell system.

The relatively poor sensitivity of P cells to low-contrast stimuli is not well understood, but it may be partially the result of two factors: receptive-field centers receive input from a very small region of the retina, and the inputs from different cone classes are subtracted from each other rather than added together. Whatever the reason, the poor performance of the P pathway in detecting low contrast implies an important role for M cells in the detection of the form of objects (which is robust at low contrasts), in addition to their role in the motion pathway (discussed later).

Summary

The retina, part of the central nervous system, is a self-contained neuronal circuit whose anatomy and physiology have been studied in great detail. There are five types of cells in the retina. Photoreceptors, bipolar cells, and ganglion cells constitute the direct, feed-forward pathway. Horizontal cells and amacrine cells subserve lateral interactions within the retina. Photoreceptors are all hyperpolarized by light, but bipolar cells can be either hyperpolarized (*off* cells) or depolarized by light (*on* cells). There is a great variety of retinal ganglion cells, but most have an antagonistic center-surround organization. In the primate, the two best studied classes of ganglion cells are P cells and M cells, which project to the parvocellular and magnocellular layers of the LGN, respectively. P cells have small receptive fields, are sensitive to the color of a stimulus, and are relatively insensitive to low contrasts. M cells have larger receptive fields, are insensitive to color, and are very sensitive to low contrasts.

THE RETINOGENICULOCORTICAL PATHWAY

Visual Information Is Relayed to the Cortex via the Lateral Geniculate Nucleus

In mammals with forward-facing eyes, such as most carnivores and the primates, retinal axons are routed so that visual information from the same points in space coming from the two eyes can be combined. From both eyes, ganglion cells whose receptive fields are in one half of the visual field project to the opposite cerebral

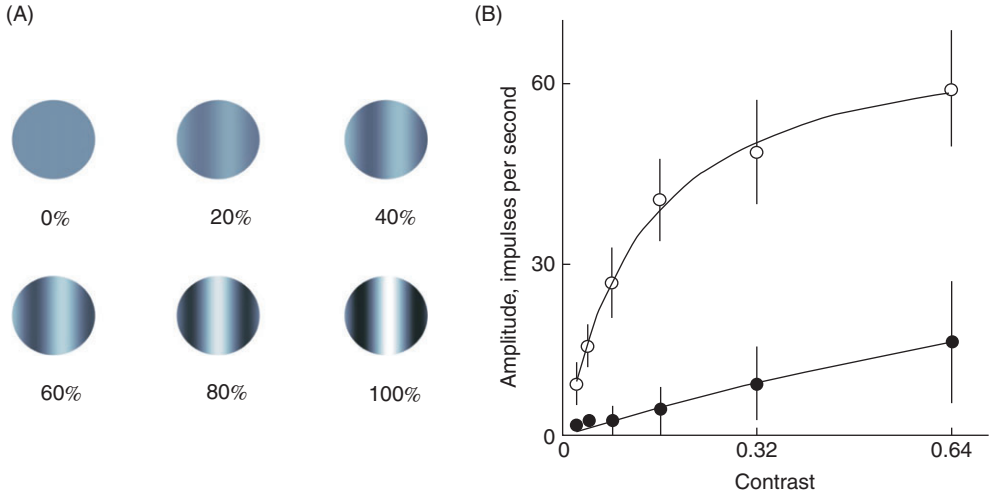


FIGURE 26.5 Relationship between stimulus contrast and the visual responses of magnocellular and parvocellular neurons in the macaque. (A) Examples of sinusoidal gratings with different levels of luminance contrast. (B) Average contrast response functions of P (•) and M (○) cells measured in spikes per second. P cells respond poorly to low contrasts and do not saturate at high contrasts. M cells respond better to low contrasts, but saturate by 20 to 30%. B: *From Kaplan and Shapley (1986).*

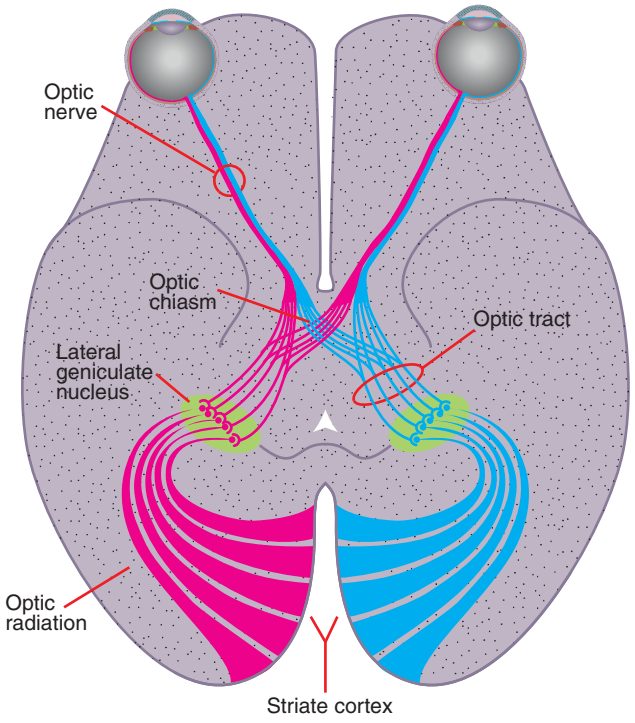


FIGURE 26.6 The retino-geniculo-cortical pathway in the human. Optic nerve axons from the nasal retina cross at the optic chiasm and join axons from the temporal retina of the other eye. Together, these contralateral and ipsilateral axons make up the optic tract, which projects to the LGN. Each of the six layers of the LGN receives input from only one eye. Axons from the LGN make up the optic radiations, which project to the striate cortex. *Adapted from Polyak (1941).*

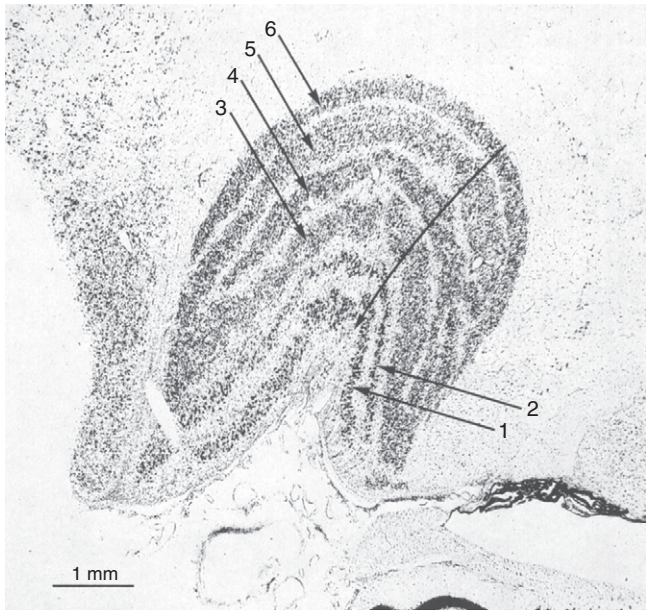


FIGURE 26.7 The six-layered LGN of the macaque monkey. The top four parvocellular layers (6, 5, 4, and 3) receive input from the ispilateral and contralateral eye in the order of contra, ipsi, contra, and ipsi. The bottom two magnocellular layers (2 and 1) receive ipsi and contra input, respectively. In between these principal layers are intercalated or koniocellular layers. The arrow from layer 6 to 1 indicates organization of the precisely aligned retino-topic maps of the six layers. The receptive fields of neurons found along this line are located at the same position in visual space. *From Hubel and Wiesel (1977).*

hemisphere (Fig. 26.6). This cross-routing occurs at the optic chiasm (from the Greek letter *chi*, χ). Here, axons from the medial (nasal, or nearer the nose) half of one retina cross over to join the axons from the lateral (temporal, or nearer the temples) half of the other retina. In other words, the temporal retina projects to the *ipsilateral* (same-side) hemisphere and the nasal retina projects to the *contralateral* (opposite-side) hemisphere. Unlike somatic sensation, which is entirely crossed (see Chapter 24), only half of the retinal axons cross. Like somatic sensation, however, the crossing of retinal axons results in each half of the external visual field being represented in the opposite cerebral hemisphere.

The LGN Is a Layered Structure That Receives Segregated Input from the Two Eyes

As they leave the optic chiasm, retinal axons from the two eyes travel together in the optic tract, which terminates in the LGN (Fig. 26.6). The LGN is composed of layers, each of which receives input from only one eye, although the number varies between species. In the macaque, the six principal layers contain most of the relay cells to primary visual cortex (Fig. 26.7). The four parvocellular layers are found dorsally. They receive P cell input from the contra-lateral and ipsilateral eye in the order of contra, ipsi, contra, and ipsi. The magnocellular layers are found more ventrally. They receive M cell input in the order of ipsi and contra. Interposed between these principal layers are the intercalated layers, populated by koniocellular neurons.

In the cat, the LGN has only two principal layers, called A and A1, which receive input from the contralateral and ipsilateral retinas, respectively. The neurons are again quite similar to either the retinal X cells or the Y cells that innervate them.

The Primary Visual Cortex in the Cat Is an Example of a Functional Hierarchy

Cat primary visual cortex (area 17) is perhaps the best understood neocortical area in any species. The first studies of Hubel and Wiesel (1962) form the foundation of this understanding. Hubel and Wiesel found cells whose responses differed dramatically from those in the retina and thalamus. Many of these cells could be excited by stimuli presented to either eye. The primary visual cortex thus represents the first level of the visual system at which binocular interactions could form a substrate for depth perception. Most strikingly, Hubel and Wiesel found that

the vast majority of cortical cells respond best to elongated stimuli at a specific orientation and give no response at the orthogonal orientation. This is in sharp contrast to cells in the retina and LGN, which are not selective for orientation.

Hubel and Wiesel described two broad classes of neurons in area 17: *simple cells* and *complex cells*. The receptive fields of simple cells were segmented into oriented *on-and-off* subregions (Figs. 26.8C and 26.8D) and were therefore most sensitive to similarly oriented stimuli. In these receptive field subregions, responses were evoked by turning a light stimulus *on* or *off*, but not by both. Given these well-defined *on-and-off* subregions in simple cells, Hubel and Wiesel suggested a straightforward hierarchical model to explain orientation selectivity: neurons in the lateral geniculate nucleus whose receptive fields are arranged in a row could all converge to excite a specific simple cell (Fig. 26.8E). Although this model has generated much controversy over the ensuing years, many of its elements have been demonstrated directly (Ferster, Chung, & Wheat, 1996; Reid & Alonso, 1995).

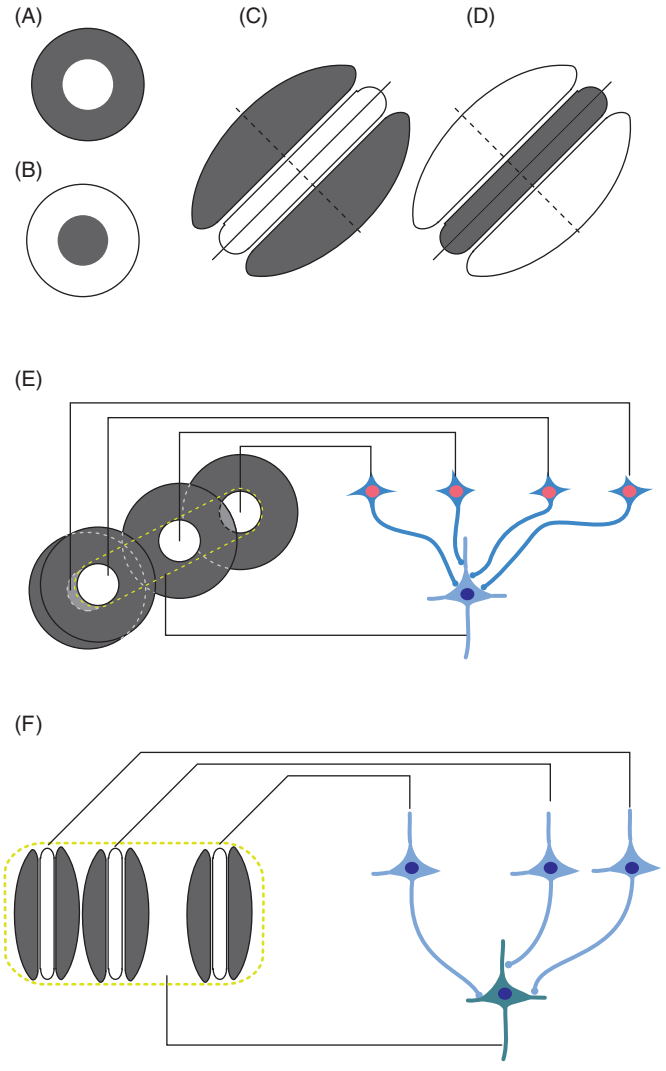


FIGURE 26.8 Hubel and Wiesel's original models of visual cortical hierarchy. (A and B) Receptive field maps of center-surround receptive fields in LGN. White: *on* responses; dark gray: *off* responses. (C and D) Receptive field maps of simple cells. (E) Model of convergent input from LGN neurons onto the cortical simple cell. (F) Model of convergent input from simple cells onto complex cells. *Adapted from Hubel and Wiesel (1962).*

The second class of cortical neurons, complex cells, also responded best to oriented stimuli, but their receptive fields were not elongated along a preferred orientation, nor were they divided into distinct *on-and-off* subregions. It would therefore be difficult to imagine how the responses of these neurons could be constructed directly from the center/surround receptive fields in the thalamus. Given the properties of simple cells, Hubel and Wiesel proposed a hierarchical scheme. Complex cells receive their orientation selectivity from convergent input from simple cells whose receptive fields have the same orientation preference, but have slightly different spatial locations (Fig. 26.8F).

This original model is hierarchical or serial; information flows from one level to the next in a well-defined series. In contrast to this hierarchical model of cortical processing, it has also been proposed that each region of the cerebral cortex is made up of several different streams of information that are all processed in parallel. As with most dichotomies, elements of both hierarchical and parallel processing have been found in the visual cortex of all mammals studied. For simplicity, an outline of the organization of visual cortex in the cat is discussed here in terms of a hierarchical model. Parallel processing will be illustrated with the example of a primate visual cortex.

The hierarchical scheme has received support from several correlations between anatomy and physiology of area 17. In broad terms, simple cells are more common in the layers of cortex that receive direct input from the thalamus (layer 4 and, to a lesser extent, layer 6). Complex cells are found more frequently in layers that are more distant from the thalamic input in layers 2 + 3, which receive input primarily from layer 4, and in layer 5, which receives most of its input from layers 2 + 3 (Gilbert & Wiesel, 1985; Fig. 26.9).

In terms of their probable function in perception, the receptive fields of complex cells furnish an early example of what was meant in the introduction by a higher order representation. Complex cells convey information about orientation to later stages of processing, but that information has been combined and generalized. A simple cell responds to an oriented stimulus of a specific configuration and a specific location; in particular, it has separate *on-and-off* subregions. A complex cell also responds to stimuli at one orientation, but the receptive field is not segregated into *on-and-off* subregions. Instead, it can respond to a light or dark stimulus of the correct orientation, independent of the exact location of the stimulus

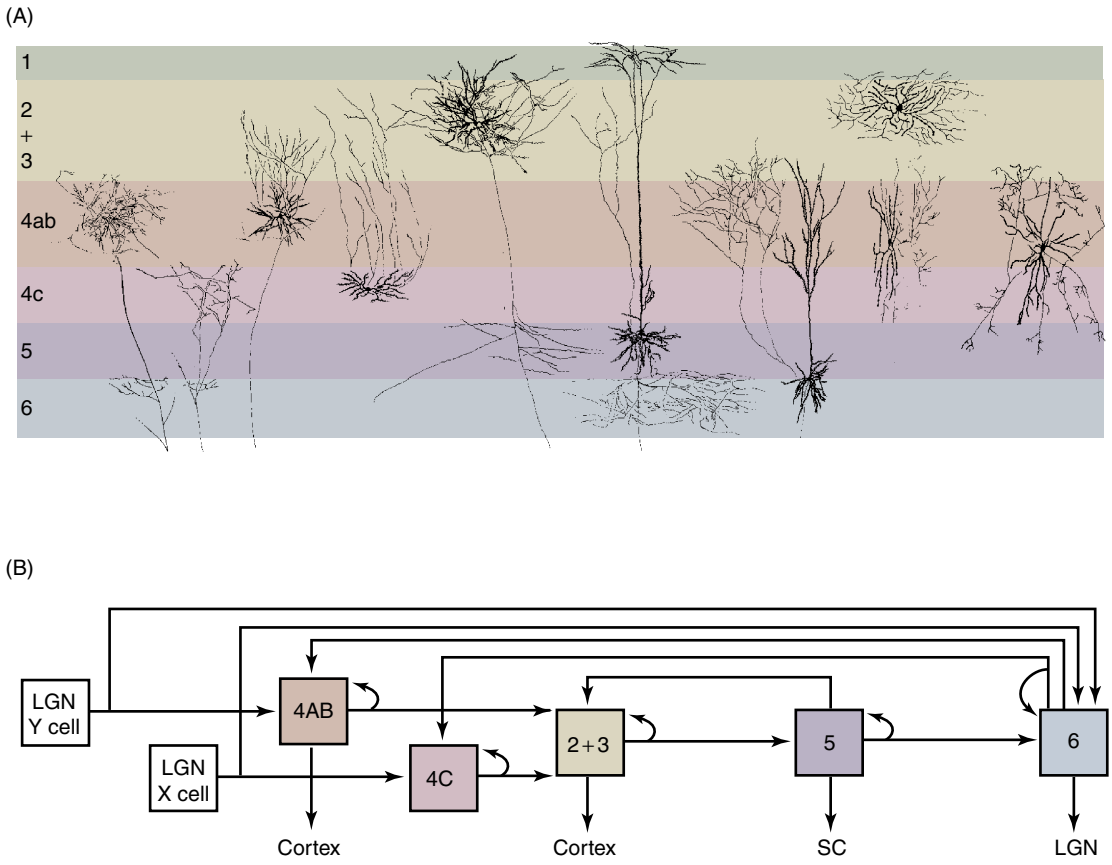


FIGURE 26.9 Circuitry of cat cortical area 17. (A) Morphology of individual cells stained with horseradish peroxidase. Thick lines: dendrites; thin lines: axons. (B) Schematic diagram of connections. Most thalamic (LGN) input is concentrated in layer 4 and, to a lesser degree, in layer 6. Y cells tend to project more superficially in layer 4 than X cells (note that sublayers 4ab and 4c are not strictly analogous to 4A, 4B, and 4C in the macaque). There is a fairly strict hierarchical pathway from layers 4 → 2 + 3 → 5 → 6 with feedback connections from 5 → 2 + 3 and 6 → 4. There are also many lateral connections between neurons within the same layer. *Adapted from Gilbert and Wiesel (1985).*

within the receptive field. This sort of generalization is a recurrent theme in the cortical processing of visual information.

Several Parallel Streams Are Found in the Macaque Primary Visual Cortex

In the macaque primary visual cortex (V1, or striate cortex), a hierarchical organization is certainly present, but it is also clearly composed of several parallel streams. As discussed earlier, at least three types of inputs to visual cortex (parvocellular, magnocellular, and koniocellular) are first segregated within the LGN. Each class of neurons projects to a specific subdivision of primary visual cortex. Most thalamic afferents terminate in layer 4C, which is split into two divisions. 4Cα receives its input from magnocellular neurons and 4Cβ from parvocellular neurons. Intercalated, or koniocellular, geniculate neurons project to layers 2 + 3, specifically to regions known as "blobs" (discussed later) that stain densely for the enzyme cytochrome oxidase (Livingstone & Hubel, 1984).

Thus, visual information from functionally and anatomically distinct M and P retinal neurons are kept separate through the LGN and at least up to the cortical neurons that receive direct thalamic input. The degree to which these pathways—as well as the more poorly understood koniocellular pathway—are kept separate within visual cortex remains an area of active research (Fig. 26.13).

Functional Architecture Can Be Seen in the Columnar Structure of the Visual Cortex

So far, the physiology of the visual cortex has been considered in terms of individual neurons and their response properties. Another major contribution of Hubel and Wiesel was their demonstration that cells with similar receptive field properties tend to be found near each other in the cortex. Further, they found that physiological response properties, such as orientation selectivity, are organized in an orderly fashion across the cortical surface. They termed the relationship between anatomy and physiology the *functional architecture* of visual cortex.

Hubel and Wiesel's key observation in this arena was that when an electrode is advanced through the cortex perpendicular to its surface, all neurons encountered have similar response properties. If a cell near the surface has a specific orientation and is dominated by input from one eye, then cells below it share these preferences. This finding is consistent with the idea, proposed by Lorente de Nó, that the fundamental unit in cortical architecture is a vertically oriented *column* of neurons. It is interesting to note that rodent cortex deviates from this organizational plan, as neurons preferring the same orientation are not clustered together in a columnar fashion (Fig. 26.14A). Hubel and Wiesel proposed that a cortical column is both a physiological and an anatomical unit, as had Mountcastle in his study of somatosensory cortex (see Chapter 24). A second observation was that the property of orientation preference varies smoothly over the cortical surface. If an electrode takes a tangential path through the cortex through many different columns, the orientation preference generally changes in a steady clockwise or counterclockwise progression, although occasionally there are discontinuous jumps (Fig. 26.10A).

In addition to orientation, at least two other parameters are mapped smoothly across the cortical surface. Cells in the visual cortex receive inputs from the two eyes in varying proportion, a property that Hubel and Wiesel termed *ocular dominance.* In tangential micro-electrode penetrations, cells are found that are dominated first by input from one eye and then the other. This provides evidence for *ocular dominance columns,* which have subsequently been demonstrated anatomically as well as physiologically. Second, prior to Hubel and Wiesel's work, a precise map of visual space across the surface of the cortex had been demonstrated. For any cortical column, receptive fields are all located at roughly the same position on the retina. Nearby columns represent nearby points in visual space in a precise and orderly arrangement. The position of a stimulus on the retina is termed its *retinotopy*; thus, a region of the brain (such as the superior colliculus, LGN, or the visual cortex) that maintains the relations between adjacent retinal regions is said to have a *retinotopic map.*

Given the existence of multiple functional maps, the obvious question is: How do these maps relate to each other? This question has been answered most definitively for the relationship between retinotopy and ocular dominance. Layer 4 in the primate is ideal for studying this question, as the receptive fields are quite small and the borders between ocular dominance columns are well defined. By making long, tangential penetrations through layer 4 of the striate cortex, Hubel and Wiesel found a precise interdigitating map from each eye. When eye dominance shifts, the receptive field location shifts to a point corresponding to the middle of the previous ocular dominance column (Fig. 26.10B). Thus, there is a 50% overlap in the spatial locations represented by adjacent ocular dominance columns. In this manner, a complete representation of space is attained for both eyes while ensuring that cells that respond to overlapping points in visual space are always nearby within the cortex.

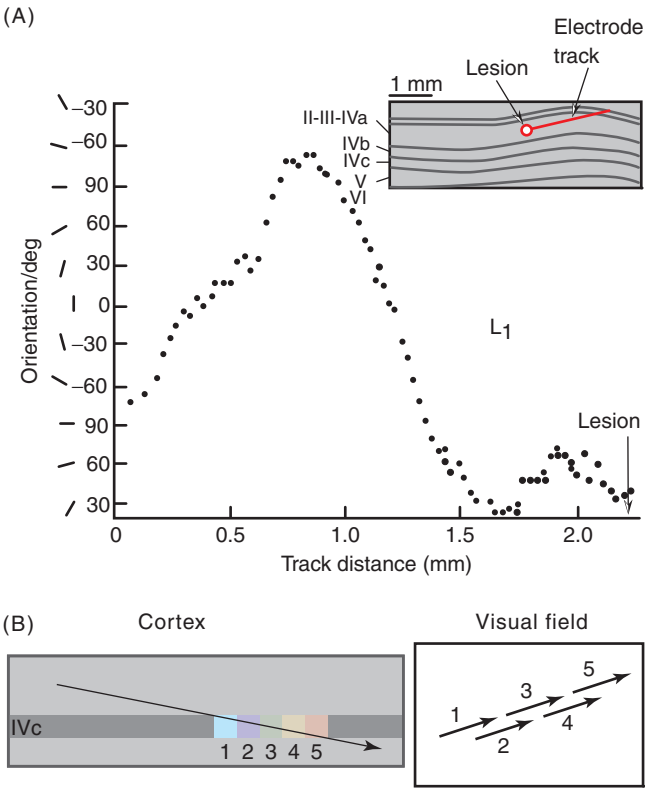


FIGURE 26.10 Two aspects of the functional architecture of the macaque primary visual cortex. (A) Graph of the preferred orientation of neurons encountered in a long microelectrode penetration through layers 2 + 3 (inset). There was a steady, slow progression of preferred orientations, although there were a few positions where the orientations changed more abruptly. (B) Schematic diagram of an electrode penetration through layer 4C (left) and the retinotopic positions of receptive fields (right). Numbers (1–5) indicate regions dominated by input from alternating eyes. At the border between ocular dominance columns (e.g., instance between 1 and 2), the location of receptive fields jumps back to a point represented near the middle of the previous ocular dominance column. There is a complete representation of visual space in columns dominated by each eye (1, 3, 5 or 2, 4) and these representations are spatially interleaved. *Adapted from Hubel and Wiesel (1977).*

Ocular Dominance and Orientation Columns Can Be Revealed with Optical Imaging

The technique of optical imaging has proven extremely useful in the study of the functional architecture of the visual cortex. Optical imaging allows direct visualization of the relative activity of small cortical ensembles rather than relying on inferences made from single unit studies (Blasdel & Salama, 1986; Grinvald, Lieke, Frostig, Gilbert, & Wiesel, 1986). The technique uses dyes that change their optical properties with neural activity. Even if no dyes are used, brain activity can be mapped with

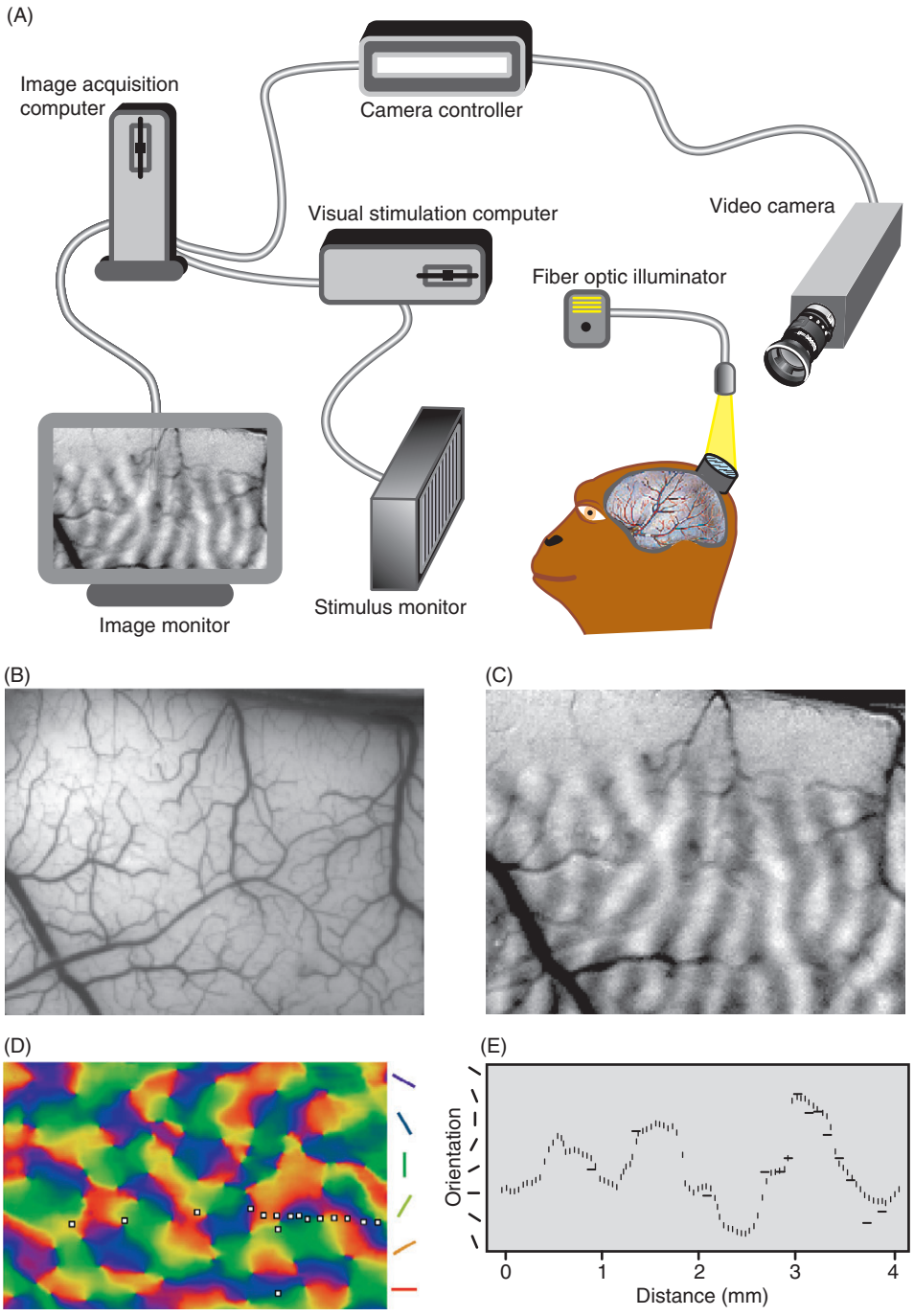


FIGURE 26.11 Optical imaging of functional architecture in the primate visual cortex. (A) Schematic diagram of the experimental setup for optical imaging. Digitized images of a region of visual cortex (as in B) are taken with a CCD camera while the anesthetized, paralyzed animal is viewing a visual stimulus. These images are stored on a second computer for further analysis. (B) Individual image (9 by 6 mm) of a region of V1 and a portion of V2 taken with a special filter so that blood vessels stand out. (C) Ocular dominance map. Images of the brain during right-eye stimulation were subtracted digitally from images taken during left-eye stimulation. (D) Orientation map. Images of the brain were taken during stimulation at 12 different angles. The orientation of stimuli that produced the strongest signal at each pixel is color coded, as indicated to the right. The key at the right gives the correspondence between color and the optimal orientation. (E) Comparison of the preferred orientation of single neurons with the optical image. At each of the locations indicated by squares in D, single neurons were recorded with microelectrodes. The preferred orientations of the neurons (dashes) were compared with the preferred orientations measured in the optical image, sampled along the line connecting the recording sites (dots). A: *Adapted from Grinvald et al. (1988)*; B and C: *Adapted from Ts'o, Frostig, Lieke, and Grinvalid (1990)*; D and E: *Adapted from Blasdel and Salama (1986)*.

an intrinsic signal, caused primarily by changes in blood flow and blood oxygenation.

A typical optical imaging experiment works as follows (Fig. 26.11A). First, a series of digitized images is taken of a region of visual cortex (Fig. 26.11B) while the animal is presented with visual stimuli through the left eye. Next, a similar series is captured during right eye stimulation. When the right eye images are subtracted digitally from the left eye images, a striking picture is created that reveals the functional architecture of ocular dominance (Fig. 26.11C). Previously, anatomical methods had been used to produce maps of ocular dominance, and these maps appear similar to those found with optical imaging, but an important feature of optical imaging is that multiple images can be obtained from the same region of the cortex. For instance, in addition to ocular dominance, maps of the preferred orientation can also be made. A useful way to present these data is in terms of a color code, in which each color represents a different preferred orientation (Fig. 26.11D). The orientation columns revealed by

optical imaging were consistent with earlier microelectrode studies (compare Fig. 26.11E with Fig. 26.10A), but the images for the first time gave a detailed picture of the layout of orientation across the cortical surface. Details of these orientation maps—and their relationships with ocular dominance—were entirely new and had not been demonstrated with previous techniques.

Cytochrome Oxidase Staining Reveals Blobs and Stripes in Cortical Areas V1 and V2

In the primate, there is a fourth feature of the functional architecture of visual cortex in addition to retinotopy, orientation, and ocular dominance columns. When stained for the enzyme cytochrome oxidase (a metabolic enzyme whose presence indicates high activity), histological sections of V1 revealed a regular pattern of patches, or blobs (Fig. 26.12). In layers 2+3, these blobs were reported to contain neurons whose receptive fields are color selective, poorly oriented, and monocular (Livingstone & Hubel, 1984).

In V2, the second visual area in the cerebral cortex (see later), cytochrome oxidase staining reveals regions of high and low activity arranged in parallel stripes (thick, thin, and pale stripes; Fig. 26.12, along the top). The anatomical connections between V1 and V2 are strongly constrained by the subdivisions revealed by cytochrome oxidase staining. In an early view, V1 neurons in the blobs projected to the thin stripes of V2, neurons in the interblobs projected to the pale stripes, and neurons in layer 4B projected to the thick stripes (Livingstone & Hubel, 1984). More recently, this organizational scheme has been revised and simplified such that V1 neurons in blob (or patch) columns, including underlying regions of layer 4B, project to the thin stripes of V2, while neurons in the interblob (interpatch) columns, again including underlying regions of layer 4B, project to both the pale stripes and thick stripes (Sincich & Horton, 2005). One of the reasons it is important to understand the anatomical relationship between V1 and V2 is that neurons in the thick, thin, and pale stripes of V2 give rise to projections that appear selective for either the dorsal or ventral streams of visual processing in extrastriate cortex (described below), whereby neurons in the thick stripes provide input to the dorsal stream and neurons in the thin and pale stripes provide input to the ventral stream.

Many Extrastriate Visual Areas Perform Different Functions

In early studies of the cytoarchitecture of cerebral cortex, the visual cortex was divided into three areas according to Brodmann's classification: areas 17, 18, and 19. These divisions were based on differences in cell cytoarchitecture, or differences in the size, morphology, and distributions of cells within the six cortical laminae. Area 17 is primary visual cortex, or striate cortex, so named because of a heavily myelinated sublamina within layer 4 (4Cα), prominently visible as a stripe in transverse section. Areas 18 and 19 were known simply as the visual association cortex. An assumption behind these designations was that areas defined by anatomical criteria would ultimately prove to be functionally specialized.

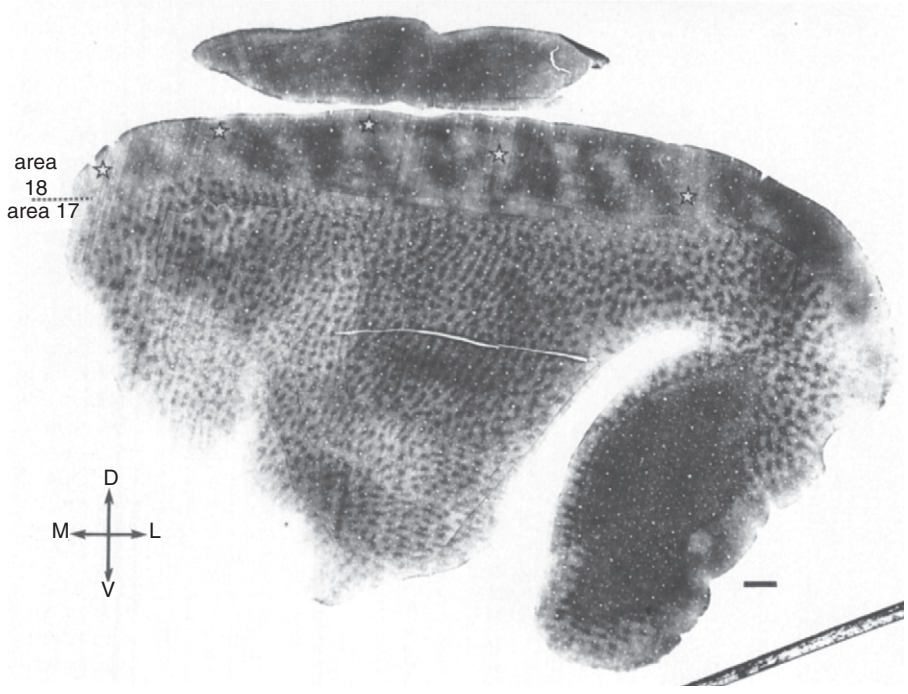


FIGURE 26.12 Macaque visual cortex stained for the metabolic enzyme cytochrome oxidase. A tangential section through layers 2 + 3 of V1 (below) and V2 (above) is shown. In V1, blobs are seen in a regular array with a spacing of 400 μm. In V2, there are cytochrome oxidase-rich stripes with much coarser spacing. The distinction between thick and thin stripes (see text) is seen less readily in the macaque than in the owl monkey. Scale bar: 1 mm. *From Livingstone and Hubel (1984).*

The demonstration of multiple visual cortical areas has been one of the important discoveries of the past quarter century in the field of sensory neurobiology. A vast expanse of cerebral cortex—greater than 50% of the total in many primate species—is involved primarily or exclusively in the processing of visual information. The extrastriate cortex now includes areas 18 and 19, as well as large regions of the temporal and parietal lobes (see Fig. 26.13 for a diagram of the four main regions of cerebral cortex: occipital, parietal, temporal, and frontal). It is composed of some 30 subdivisions that can be distinguished by their physiology, cytoarchitecture, histochemistry, and/or connections with other areas (Felleman & Van Essen, 1991). Each of these extrastriate visual areas is thought to make unique functional contributions to visual perception and visually guided behavior. Higher-order cortical areas are also present in rodents (Fig. 26.14B); however, the number of cortical areas serving vision in rodents is substantially less in comparison to primates.

As a testament to the increasing complexity of the field, the naming of visual areas (Fig. 26.13) has progressed from the use of simple labels—V1 through V4—to the use of more complex terms specifying anatomical location of each new area, such as MT (medial temporal, or V5), MST (medial superior temporal), or IT (inferotemporal, itself made up of several distinct areas).

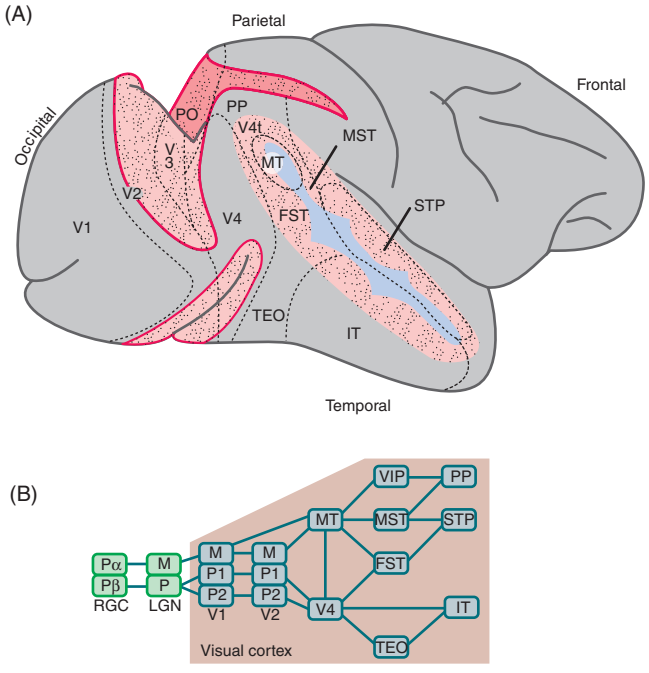


FIGURE 26.13 Extrastriate cortical regions. (A) Lateral view of the macaque brain, with the sulci partially opened to expose the areas within them. Shown are the rough outlines of the main visual areas, which take up all of the occipital cortex and much of the parietal and temporal cortex. (B) Partial diagram of the connections between visual areas. Emphasis is placed on the hierarchical organization of the connections and on the partially segregated P parvocellular and M magnocellular pathways. *Adapted from Albright (1993).*

The dual themes of parallel and hierarchical processing are central to the understanding of the extrastriate cortex. Figure 26.13 is a vastly simplified version of a wiring diagram between visual areas (Felleman & Van Essen, 1991). Several criteria have been used to define new visual cortical areas. First, extrastriate regions have retinotopic maps of the visual world that can be demonstrated by physiological recordings. Second, a clear hierarchy between many cortical areas can be demonstrated anatomically. There is a stereotyped pattern of projections from one visual area to the next. Where strong connections between two areas exist, these connections tend to be bidirectional. The feedforward and feedback connections are distinguished by the layers that send and receive the connections. In such a manner, a clear hierarchy can be traced, for instance, along the pathway V1 → V2 → V3 → MT → MST (with several shortcuts, such as V1 → MT).

While physiological studies of cortical areas revealed profound differences between them, a series of studies by Ungerleider and Mishkin (1982) uncovered a higher-order dichotomy between two types of processing in the extrastriate cortex. Using behavioral analyses of animals with anatomically defined cortical lesions, Ungerleider and Mishkin found a strong dissociation between the types of deficits exhibited by animals with lesions in either their parietal cortex or their temporal cortex (see Fig. 26.13). Animals with temporal lesions were often much worse at recognizing objects visually, although lower-level visual function, such as acuity, was not appreciably lessened. Parietal lesions led to little or no deficit in object recognition, but visuospatial tasks, such as visually guided behavior, were impaired profoundly. From these studies and the growing body of physiological evidence, two distinct streams were postulated: a *temporal* or *ventral stream* devoted to object recognition and a *parietal* or *dorsal stream* devoted to action or spatial tasks. Ungerleider and Mishkin termed these the "what" and "where" pathways.

The ventral stream, V1 → V2 → V3 → V4 → IT …, is discussed elsewhere (Chapter 44). This chapter considers only the dorsal stream, V1 → V2 → MT → MST…. The dorsal stream is dominated by magnocellular cells and the ventral stream by parvocellular cells, although the segregation is far from strict (Merigan & Maunsell, 1994).

In the Parietal Cortex, Neurons Are Selective for Higher Order Motion Cues

MT, or V5, is perhaps the best understood of the extrastriate visual areas. It has held particular interest ever since it was discovered in 1971 (Allman & Kaas, 1971; Dubner & Zeki, 1971), primarily because it was the first area found that was strongly dominated by one visual

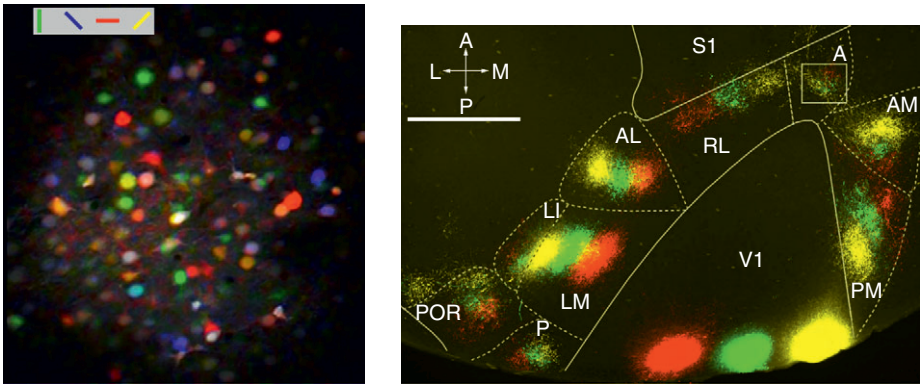


FIGURE 26.14 Rodent visual cortex. (A) Orientation-selective neurons in rat visual cortex (as well as mice) are not organized into columns, but instead are positioned randomly in V1. Individual neurons were stained with an activity-sensitive indicator and imaged *in vivo* with two-photon microscopy, a form of fluorescence imaging that can easily resolve single neurons in the living brain. Neurons were imaged during visual stimulation and their preferred orientations are indicated with a color code, as in Fig. 26.12C (legend serves also as 100 μm scale bar). (B) Anatomical image of primary visual cortex (V1) and nine distinct higher visual areas in the mouse. Higher visual areas are named according to their anatomical positions, such as posteromedial (AM) or anterolateral (AL); S1 is primary somatosensory cortex. Three different colors of fluorescent tracers were injected in V1, so that axons from V1 to each of the higher visual areas could be visualized. Axons from the three distinct injection sites project to distinct regions within each of the higher areas, which indicates that the retinotopic map of visual space is maintained in all of the visual areas. Scale bar: 1 mm. Legend: **A**nterior, **P**osterior, **L**ateral, **M**edial. A: *Adapted from Ohki, Chung, Ch'ng, Kara, and Reid (2005)*; B: *Adapted from Ma and Burkhalter (2007).*

function. Fully 95% of the neurons in MT are highly selective for the direction of motion of a stimulus (Dubner & Zeki, 1971). In V1, a significant fraction of neurons are selective for the direction of motion, but the optimal speed may vary depending on the spatial structure of the object that is moving. In MT, speed tuning is less dependent on other stimulus attributes. Receptive fields of individual neurons in MT integrate motion information over large regions of visual space and are qualitatively less selective than neurons in the primary visual cortex. This generalization of motion signals can be achieved in a simple manner, such as by adding together inputs over space or, in a complex manner, by combining two component motions in different directions into a single coherent motion. Of even greater interest, neurons in MT (and, to a greater extent, those in MST) appear to be sensitive to more complex aspects of visual motion, such as the motion of extended objects rather than isolated features (Albright, 1993).

The range of stimuli that a given MT neuron can respond to is impressively broad—the attributes, or form cues, that define a figure can be luminance, texture, or relative motion—but the preferred direction and speed are always the same for that neuron (Albright, 1993). This is an extreme example of what were termed high-order responses in the introduction to this chapter. Neurons at lower levels in the visual system are sensitive to isolated and specific features in visual scenes. Higher visual areas respond to very specific attributes, but these attributes are increasingly remote from the physical stimulus. Instead, they represent increasingly complex concepts, such as the motion of an extended object or the identity of a face.

Summary

The lateral geniculate nucleus of the thalamus is a layered structure that receives segregated input from the two eyes and projects to the primary visual cortex (V1). Geniculate neurons have center-surround receptive fields that are similar to those of their retinal inputs. In contrast, most neurons in the primary visual cortex are sensitive to the orientation of a stimulus. In mammalian species, there are elements of both hierarchical and parallel processing in the primary visual cortex. This section has emphasized that the visual cortical circuit in the cat can be seen as a functional hierarchy that transforms geniculate input into simple and then complex receptive fields. In the macaque monkey the focus has been on the parallel pathways that are kept relatively separate in the striate and extrastriate cortex.

The visual cortex has an orderly functional architecture. Neurons within a cortical column have similar receptive field attributes, such as orientation selectivity, ocular dominance, and receptive field location. Each of these receptive field attributes varies smoothly over the cortical surface. The organization of ocular dominance columns and orientation columns across the cortical surface can be visualized with optical imaging.

Finally, in the macaque monkey, there are more than 30 extrastriate visual cortical regions, each of which performs different functions. These extrastriate regions can be divided into two pathways: the ventral stream, devoted to form recognition, and the dorsal stream, devoted to action or to spatial tasks. Areas in each stream form a

functional and anatomical hierarchy. Neurons in successive visual areas respond to increasingly high-order or abstract features of the visual world.

References

Albright, T. D. (1993). Cortical processing of visual motion. In F. A. Miles & J. Wallman (Eds.), *Visual motion and its role in the stabilization of gaze* (pp. 177–201). Amsterdam: Elsevier Science.

Allman, J. M., & Kaas, J. H. (1971). A representation of the visual field in the caudal third of the middle temporal gyrus of the owl monkey *(Aotus Trivirgatus). Brain Research, 31,* 85–105.

Berson, D. M. (2003). Strange vision: Ganglion cells as circadian photoreceptors. *Trends in Neuroscience, 26,* 314–320.

Blasdel, G. G., & Salama, G. (1986). Voltage-sensitive dyes reveal a modular organization in monkey striate cortex. *Nature, 321,* 579–585.

Chalupa, L. M., & Williams, R. W. (Eds.). (2008). *Eye, retina, and visual system of the mouse*. Cambridge, MA: MIT Press.

Dacey, D. M. (2000). Parallel pathways for spectral coding in primate retina. *Annual Review of Neuroscience, 23,* 743–775.

Dowling, J. E. (1997). Retina. In *Encyclopedia of human biology* (2nd ed., vol. 7, pp. 571–587). Ed. R Dulbecco. New York, NY: Academic Press.

Dubner, R., & Zeki, S. M. (1971). Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. *Brain Research, 35,* 528–532.

Enroth-Cugell, C., & Robson, J. G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *Journal of Physiology, 187,* 517–552.

Felleman, D. J., & van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1,* 1–47.

Ferster, D., Chung, S., & Wheat, H. (1996). Orientation selectivity of thalamic input to simple cells of cat visual cortex. *Nature, 380,* 249–252.

Gilbert, C. D., & Wiesel, T. N. (1985). Intrinsic connectivity and receptive field properties in visual cortex. *Vision Research, 25,* 365–374.

Grinvald, A., Lieke, E., Frostig, R. D., Gilbert, C. D., & Wiesel, T. N. (1986). Functional architecture of cortex revealed by optical imaging of intrinsic signals. *Nature, 324,* 361–364.

Helmholtz, H. (2000). In J. P. C. Southall (Ed.), *Handbook of physiological optics* (vol. 2, pp. 143). Bristol, UK: Thoemmes Press.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology, 160,* 106–154.

Kaplan, E., & Shapley, R. M. (1986). The primate retina contains two types of ganglion cells, with high and low contrast sensitivity. *Proceedings of the National Academy of Sciences of the United States of America, 83,* 2755–2757.

Kim, I. J., Zhang, Y., Yamagata, M., Meister, M., & Sanes, J. R. (2008). Molecular identification of a retinal cell type that responds to upward motion. *Nature, 452,* 478–482.

Kuffler, S. W. (1953). Discharge patterns and functional organization of the mammalian retina. *Journal of Neurophysiology, 16,* 37–68.

Livingstone, M. S., & Hubel, D. H. (1984). Anatomy and physiology of a color system in the primate visual cortex. *Journal of Neuroscience, 4,* 309–356.

Merigan, W. H., & Maunsell, J. H. R. (1994). How parallel are the primate visual pathways? *Annual Review of Neuroscience, 16,* 369–402.

Movshon, J. A., Adelson, E. H., Gizzi, M., & Newsome, W. T. (1985). The analysis of moving visual patterns. In C. Chagas, R. Gattass, & C. G. Gross (Eds.), *Pattern recognition mechanisms* (pp. 117–151). Rome: Vatican Press.

Ohki, K., Chung, S., Ch'ng, Y. H., Kara, P., & Reid, R. C. (2005). Functional imaging with cellular resolution reveals precise microarchitecture in visual cortex. *Nature, 433,* 597–603.

Ratliff, F. (1974). *Studies on excitation and inhibition in the retina: A collection of papers from the laboratories of H. Keffer Hartline*. New York: Rockefeller Univ. Press.

Reid, R. C., & Alonso, J. M. (1995). Specificity of monosynaptic connections from thalamus to visual cortex. *Nature, 378,* 281–284.

Sincich, L. C., & Horton, J. C. (2005). The circuitry of V1 and V2: Integration of color, form, and motion. *Annual Review of Neuroscience, 28,* 303–326.

Ts'o, D. Y., Frostig, R. D., Lieke, E. E., & Grinvald, A. (1990). Functional organization of primate visual cortex revealed by high resolution optical imaging. *Science, 249,* 417–420.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.

Wiesel, T. N., & Hubel, D. H. (1966). Spatial and chromatic interactions in the lateral geniculate body of the rhesus monkey. *Journal of Neurophysiology, 29,* 1115–1156.

Wu, S. M. (1994). Synaptic transmission in the outer retina. *Annual Review of Physiology, 56,* 141–168.

Wang, Q., & Burkhalter, A. (2007). Area map of mouse visual cortex. *Journal of Comparative Neurology, 502,* 339–357.

Suggested Readings

Boynton, R. M., & Kaiser, P. K. (1996). *Human color vision*. (2nd ed.). Washington, DC: Optical Society of America.

Chalupa, L. M., & Werner, J. S. (Eds.). (2004). *The visual neurosciences*. Cambridge, MA: MIT Press.

Hubel, D. H. (1988). *Eye, brain, and vision*. New York, NY: Freeman.

Hubel, D. H., & Wiesel, T. N. (2005). *Brain and visual perception*. New York, NY: Oxford University Press.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York, NY: Freeman.

Rodieck, R. W. (1998). *The first steps in seeing*. Sunderland, MA: Sinauer Associates.

Wandell, B. A. (1993). *Foundations of vision*. Sunderland, MA: Sinauer.

R. Clay Reid and W. Martin Usrey

This page intentionally left blank

SECTION V

MOTOR SYSTEMS

This page intentionally left blank

CHAPTER

27

Fundamentals of Motor Systems

"All mankind can do is to move things … whether whispering a syllable or felling a forest."—Sherrington

This quote is a reminder of the basic fact that all interactions with the surrounding world are through the actions of the motor system. When a human baby is born, it is a sweet but very immature survival machine, with a limited behavioral repertoire. It is able to breathe and has searching and sucking reflexes so that it can be fed from the mother's breast. It can swallow, vomit, and process food, and it can cry for attention if something is wrong. A baby also has a variety of protective reflexes that mediate coughing, sneezing, and touch avoidance. These different patterns of motor behavior are thus available at birth and are due to innate motor programs (Fig. 27.1).

During roughly the first 15 years of life, the motor system continues to develop through maturation of neuronal circuitry and by learning through different motor activities. Playing represents an important element both in children and in young mammals such as kittens and pups. During the first year of life, the human infant matures progressively. It can balance its head at 2 to 3 months, is able to sit at around 6 to 7 months, and can stand with support at approximately 9 to 12 months. The coordination of different types of posture, such as standing, is a complex motor task to master, with hundreds of different muscles taking part in a coordinated fashion. Sensory information contributes importantly, in particular from the vestibular apparatus, eyes, muscle, and skin receptors located at the soles of the feet. This development represents to a large degree a maturation process following a given sequence but with individual variability among different children. When the postural system has evolved to a sufficient degree, a child is able to start walking, which requires that the body posture is maintained while the points of support are changing by the alternating movements of the two legs. In common language the child is said to "learn" to walk, but in reality a progressive maturation of the nervous system is taking place. Identical twins start to walk essentially at the same time, even if one has been subjected to training and the other has not. At this point the motor pattern is still very immature. Proper walking coordination followed by running appears later, and the basic motor pattern actually continues to develop until puberty. The fine details of the motor pattern are adapted to the surrounding world, but also to modification by will. The basic motor coordination underlying reaching and the fine control of hands and fingers undergo a similar characteristic maturation process over many years.

While the newborn human infant is comparatively immature, other mammals, such as horses and deer, represent another extreme. The gnu, an African buffalo-like antelope, needs to run away in order to survive attacks of predators such as lions. The young calf of the gnu can stand and run directly after birth and has been reported to be able to gallop ahead 10 minutes after delivery, tracking the running mother (Fig. 27.2). Clearly the neural networks underlying locomotion, equilibrium control, and steering must be sufficiently mature and available at birth, needing minimal calibration. This is astounding. A similar range of maturity is present in birds. To get out of the egg, a chick makes coordinated hatching movements to open up the eggshell to subsequently lift off the top of the egg and to stand up to walk away on two legs following the mother hen and start picking at food grains. Most birds are more immature when hatching, but after a few weeks, they leave the nest flying rather successfully, for the first time in their life, and thus without any previous experience.

© 2013 Elsevier Inc. All rights reserved.

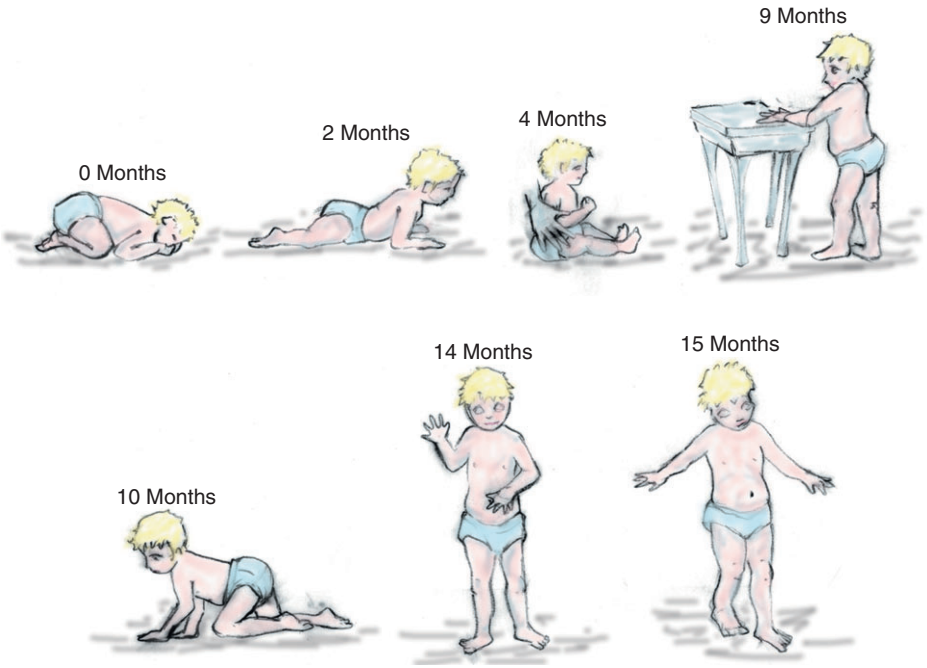


FIGURE 27.1 Motor development of the infant and young child. The pattern of maturation of the motor system follows a characteristic evolution. Two months after birth a child can lift its head, at 4 months it sits with support, and subsequently it is able to stand with support; later it crawls, stands without support, and finally walks independently. The approximate time at which a child is able to perform these different motor tasks is indicated above each figure. The variability in the maturation process is substantial. *Modified from M. M. Shirley.*

FIGURE 27.2 Some animals are comparatively mature when they are born. Ten minutes after the calf of the gnu, a buffalo-like antelope, is born, it is able to track its mother in a gallop. This means that the postural and locomotor systems are sufficiently mature to allow the young calf to generate these complex patterns of motor coordination at birth. There is little time to calibrate the motor system after birth and obviously no time for learning. *Courtesy of Erik Tallmark.*

In addition to the basic motor skills such as standing, walking, and chewing, humans also develop skilled motor coordination, allowing delicate hand and finger movements to be used in handwriting or playing an instrument or utilizing the air flow and shape of the oral cavity to produce sound as in speech or singing. The neural substrates allowing learning and execution of these complex motor sequences are expressed genetically and are characteristic of our species. What is learned, however, such as which language one speaks or the type of letters one writes, is obviously a function of the cultural environment.

BASIC COMPONENTS OF THE MOTOR SYSTEM

Motoneurons and Motor Units

The motoneurons that control different muscles are located in different motor nuclei along the spinal cord and in the brainstem. Each motoneuron sends its axon to one muscle and innervates a limited number of muscle fibers. A motoneuron with its muscle fibers is referred to as a motor unit. The muscle fibers of each motor unit have similar contractile properties and metabolic profile. The muscle fibers in different muscles are composed of three main types specialized for different demands, such as a continuous effort as in long-distance running (slow motor units) or fast explosive movements, such as lifting a heavy object [fast motor units (two subtypes)].

Motoneurons are activated by interneurons of different motor programs or reflex centers and by descending tracts from the forebrain and the brainstem. Thus motoneurons supplying different muscles can be activated with great precision by these different sources that together determine the degree of activation, as well as the exact timing of the motoneurons of a given muscle.

Sensory Receptors Are Important in Movement Control

Signals from sensory receptors are used by the motor system in a number of ways.

1. Sensory signals can trigger behaviorally meaningful motor acts such as withdrawal, coughing, or swallowing reflexes.

2. Sensory receptors may contribute to the control of an ongoing motor pattern and influence the switch from one phase of movement to another. For instance, in the case of breathing, sensory signals from lung volume receptors control when inspiration is terminated. In walking, sensory signals related to hip position and load on the limb help regulate the duration of the support and swing phases of the step cycle and correct for perturbations.
3. Sensory signals may also be more specific and influence only the level of activity of one muscle or a group of close synergists. Muscle receptors play a particularly important role in this context. They are of two types: Golgi tendon organs that sense the degree of contraction in a muscle (located in series with the muscle fibers at their insertion into the tendon), and muscle spindles that signal the length of a muscle (located in parallel to the muscle fibers), as well as the speed of changes in length. Moreover, the sensitivity of muscle spindles can be regulated actively through a separate set of small motoneurons, referred to as γ-motoneurons (in contrast to the larger α-motoneurons that control muscle contraction). These two muscle receptors have fast conducting afferent axons that provide rapid feedback to the spinal cord and take part in the autoregulation of the motor output to a given muscle. Fast muscle spindle afferents provide direct monosynaptic excitation to the α-motoneurons that control the muscle in which the muscle spindle is located. Thus, lengthening a muscle will lead to excitation of its motoneurons, counteracting the lengthening. This stretch reflex thus involves negative feedback. This reflex arc also provides the basis of the muscle contraction elicited by a brief tendon tap, which is used as a clinical test to investigate whether responsiveness of the motoneuronal pool is normal (the tendon reflex) (Fig. 27.3).
4. Sensory signals from a number of different receptor systems help detect and counteract any disturbance of body posture during standing or maintenance of another body position. Skin and muscle receptors and receptors signaling joint position, as well as vestibular receptors and vision, contribute to different aspects of the dynamic and static control of body position.
5. When an object is held by the fingers, skin receptors at the contact points at the fingertips play an important role. If the object tends to slip, the skin receptors become activated and signal rapidly to the nervous system that there is a need for extra muscle force.
6. A great variety of sensory signals provide information about the position of different parts of the body in relation to each other and to the external world. Such information is critical when initiating a voluntary movement. For example, to move a hand toward a given object, it is important to know the initial position of the arm and the hand in space. Depending on whether the hand is located to the left or the right of the object, different types of motor commands must be given to bring the hand to the target. Sensory information about the initial position of the hand (or any other body part) in relation to the rest of the body, as well as the target for the movement, is thus important for eliciting an effective neural command signal.

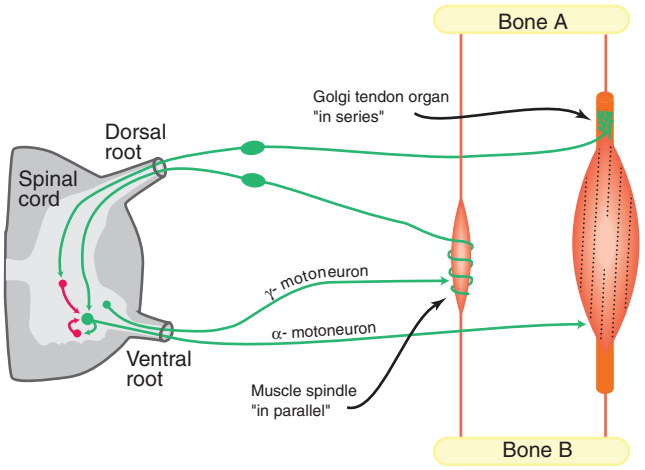


FIGURE 27.3 Feedback from muscle receptors to motoneurons: the stretch reflex. In a muscle connected between two bones, the Golgi tendon organ is located at the transition between muscle and tendon. It senses any active tension produced by the muscle fibers being in series between muscle and tendon. The muscle spindle is located in parallel with the muscle fibers. It signals the muscle length and dynamic changes in muscle length. Both the muscle spindle and the Golgi tendon organ have fast-conducting afferent nerve fibers in the range of ~100 m/s. The muscle spindle activates α-motoneurons directly, which causes the muscle fibers to contract. The sensitivity of the muscle spindle can be actively regulated by γ-motoneurons, which are more slowly conducting. The muscle spindle provides negative feedback. If the muscle with its muscle spindle is lengthened, the afferent activity from the muscle spindle increases, exciting the α-motoneurons and leading to an increased muscle contraction, which in turn counteracts the lengthening. This is called the stretch reflex. The Golgi tendon organ provides force feedback. The more the muscle contracts, the more the Golgi tendon organ and its afferent are activated. In the diagram the intercalated interneuron between afferent nerve fiber and motoneuron is inhibitory. Thus increased muscle force leads to inhibition of the α-motoneuron, which results in a decrease of muscle force. The efficacy of length and force feedback can be regulated independently in the spinal cord and via γ-motoneurons. Thus their respective contributions can vary considerably between different patterns of motor behavior.

To summarize, the sensory contribution to motor control is very important in many different contexts. If sensory control is incapacitated, motor performance will in most cases be degraded. Without sensory information in different forms, movements can usually still be executed, but as a rule with much less perfection.

Feedback and Feed Forward Control

If a perturbation occurs when an object is held in the hand, or when standing in a moving bus, it will be detected by different sensory receptors. This sensory signal will be fed back to the nervous system and be used to counteract the perturbation rapidly. This represents feedback control that is a correction of the actual perturbation after it has occurred. A limiting factor for the efficacy of feedback control in biological systems is the delay involved. A sensory afferent signal must first be elicited in the receptors concerned. It then has to be conducted to the central nervous system and be processed there to determine the proper response. The correction signal must subsequently be sent back to the appropriate muscle(s) and make the muscle fibers build up the contractile force required. In large animals, including humans, the delays involved can be substantial, and during fast movement sequences there may be little or no time for feedback corrections. In less demanding, more static situations, such as standing, sensory feedback is of critical importance.

In many cases a perturbation is anticipated before it is initiated, and correction begins before it has actually occurred. This type of proactive control is often called feed forward control in contrast to feedback. Such anticipatory control mechanisms are often automatic and involuntary, and are part of an inherent control strategy. For example, when standing on two legs and planning to lift up one leg to stand on the other, the body position begins to shift over to the supporting leg before the other leg is lifted. The projection of the center of gravity will fall between the two feet initially and will have to shift over to a projection through the supporting limb. In most instances, movement commands are designed so that the corrections of body posture required for a stable movement occur before the particular movement has started, as when lifting a heavy object. If the converse situation occurred, and body position was corrected only after a perturbation had taken place, movements would become much less precise than with feed forward control.

MOTOR PROGRAMS COORDINATE BASIC MOTOR PATTERNS

Both vertebrates and invertebrates have preformed microcircuits—neuronal networks that contain the necessary information to coordinate a specific motor pattern such as swallowing, walking, or breathing. When a given neuronal network is activated, the particular behavior it controls will be expressed. A typical network consists of a group of interneurons that activate specific groups of motoneurons in a certain sequence and inhibit other motoneurons that may counteract the intended movement. Such a group of interneurons is often referred to as a central pattern generator (CPG) or motor program. This CPG can be activated by will, as at the start of walking, or be triggered by sensory stimuli, as in a protective reflex or swallowing. In most types of motor behavior, sensory feedback may also form an integral part of the motor control circuit and may determine the duration of the motor activity such as the inspiratory phase of breathing (Fig. 27.4).

There are several different types of motor programs that are located along the neuraxis from the spinal cord to the forebrain. Together they constitute a set of standard but adaptable motor programs that can be referred to as a motor infrastructure (Fig. 27.5A).

The intrinsic function of the neuronal networks that comprise the CPGs have in most cases been difficult to

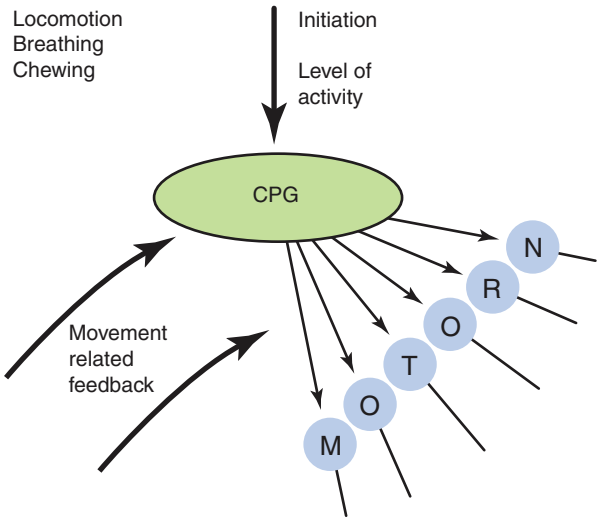


FIGURE 27.4 Motor coordination through interneuronal networks: central pattern generators. The brainstem and spinal cord contain a number of networks that are designed to control different basic patterns of the motor repertoire, such as breathing, walking, chewing, or swallowing. These networks are often referred to as central pattern generator networks (CPGs). CPGs contain the necessary information to activate different motoneurons and muscles in the appropriate sequence. Some CPGs are active under resting conditions, such as that for breathing, but most are actively turned on from the brainstem or the frontal lobes. For instance, the CPG coordinating locomotor movements is turned on from specific areas in the brainstem, referred to as locomotor centers. These descending control signals not only turn on the locomotor CPG, but also determine the level of activity in the CPG and whether slow or fast locomotor activity will occur. In order to have the motor pattern well adjusted to external conditions and different perturbations, sensory feedback acts on the CPG and can modify the duration of different phases of the activity cycle, providing feedback onto motoneurons. Although vertebrates as well as invertebrates have CPGs for a great variety of motor functions, the intrinsic operation of these networks of interneurons constituting the CPG is unclear in most cases. In invertebrates, a few CPG networks have been studied in great detail, such as the stomatogastric system of the lobster and networks coordinating the activity of the heart and locomotion in the leech. In vertebrates, detailed information concerning locomotor networks is only available in lower forms, such as the frog embryo and the lamprey.

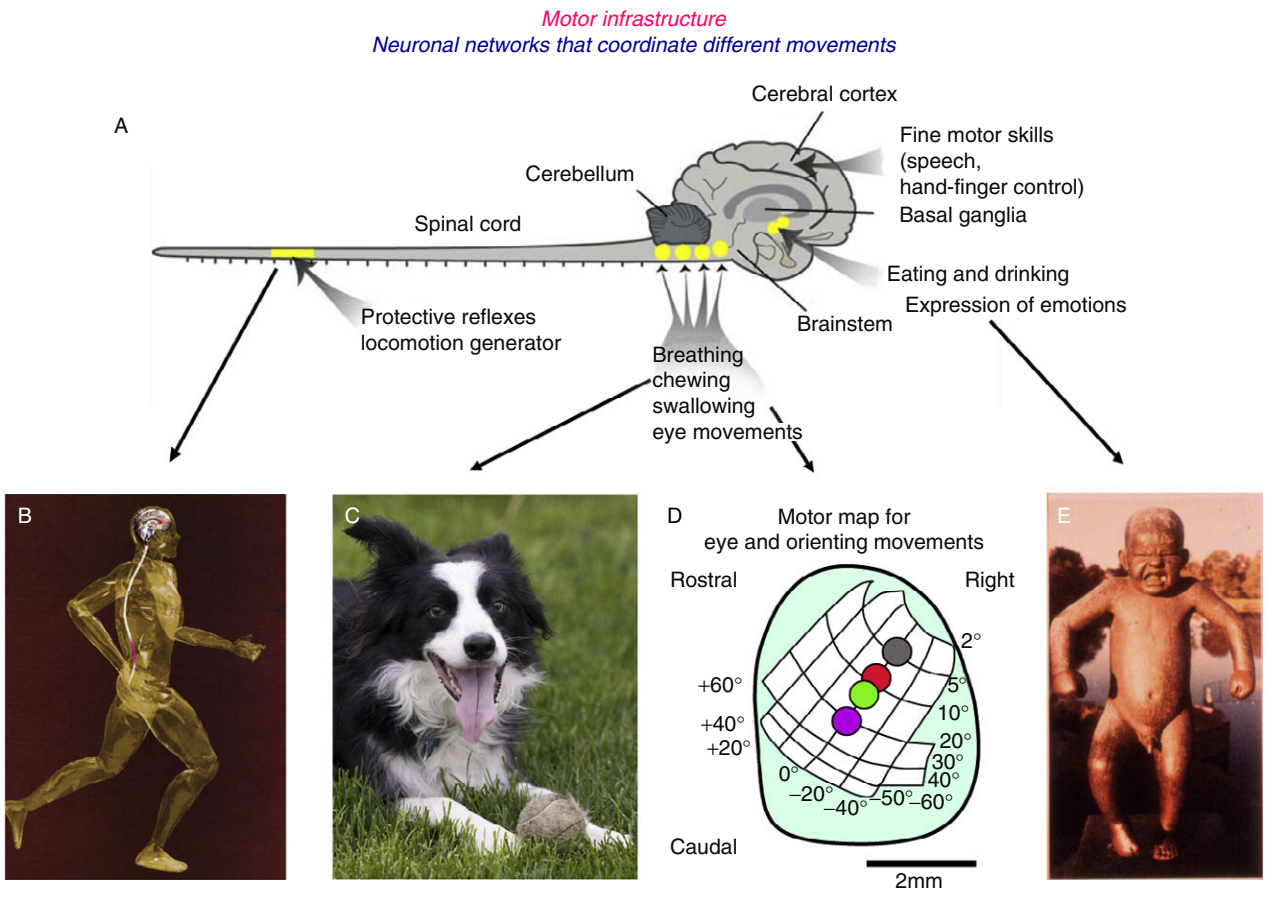


FIGURE 27.5 Location of different networks (CPGs) that coordinate different motor patterns in vertebrates. (A) Represents an overview of the location of the different CPGs in vertebrates. The spinal cord contains CPGs for locomotion (B) and protective reflexes, whereas the brainstem contains CPGs for breathing (C), chewing, swallowing, and a motor map for saccadic eye movements (D). The hypothalamus in the forebrain contains centers that regulate eating and drinking. These areas can coordinate the sequence of activation of different CPGs. For instance, if the fluid intake area is activated, the animal starts looking around for water, walks toward the water, positions itself to be able to drink, and finally starts drinking. The animal will continue to drink as long as the stimulation of the hypothalamic area is sustained. This is an example of recruitment of different CPGs in a behaviorally relevant order. The expression of emotion is represented by seven motor programs as illustrated for anger (E) by a child sculptured by Vigeland. The cerebral cortex is important, particularly for fine motor coordination involving hands and fingers and for speech.

deduce due to the complexity of the neuronal interactions. Box 27.1 provides an example from the locomotor CPG of the lamprey, in which the brainstem-spinal cord networks are understood.

Motor Patterns Triggered as a Reflex Response

1. Protective skin reflexes lead to withdrawal of the stimulated part of the body from a stimulus that may cause pain or tissue damage. The central program consists in this case of a group of interneurons that activate motoneurons and muscles, giving rise to an appropriate withdrawal movement. Motoneurons to antagonistic muscles are inhibited (reciprocal inhibition). The withdrawal reflexes of the limbs are often referred to as flexion reflexes, but in reality they represent a family of specific reflexes, each of which is designed to remove the specific skin region involved from the painful stimulus.
2. Coughing and sneezing reflexes remove an irritant from the nasal or tracheal mucosa by inducing a brief pulse of air flow at very high velocity (at storm or hurricane speeds). This is caused by an almost synchronized activation of abdominal and respiratory muscles, coordinated by a CPG, the activity of which is triggered by afferents activated by the irritant.
3. Swallowing reflexes are activated when food is brought in contact with mucosal receptors near the pharynx. This leads to a coordinated motor act with sequential activation of different muscles that propel the food bolus through the pharynx down the esophagus to the stomach. In this case the CPG is able to coordinate the motor act over several seconds, in contrast to the previous motor acts in which the activity occurs almost simultaneously.

BOX 27.1

CENTRAL PATTERN GENERATOR NETWORKS

To establish a network as a central pattern generator network (CPG), one needs to show that the motor pattern can be generated in the absence of sensory input. This is the case for all networks discussed in this chapter, although in most cases sensory information interacts with the CPG for phase transitions.

To understand the intrinsic function of a given network, one needs as a minimal requirement to identify which nerve cells are part of the network, the synaptic interactions within the network, and the membrane properties of each cell type. A network like the one controlling locomotion (Fig. B27.1) can be turned on by signals from the brainstem commands region. If a population of excitatory interneurons (E) is turned on, and they interact synaptically with one another to some degree, burst activity can be generated. Such groups of excitatory interneurons can be said to form a burst generating kernel. It is easy to understand that when a group of interneurons are activated, they will generate a train of action potentials, but what is the mechanism responsible for terminating the activity and thus for generating a burst? In the spinal network generating locomotion in the lamprey, the cellular properties of the neurons are of critical importance. During each burst calcium and sodium ions accumulate intracellularly due to both synaptic and spike activity that engages different types of ion channels (Fig. B27.1). This in turn leads to an activation of calcium-dependent K^+ channels and also potentially sodium-dependent K^+-channels. They act in a hyperpolarizing direction in each neuron and help terminate the activity. Since the neurons interact, there will also be less excitatory drive as neurons progressively drop out. In Figure B27.1, the excitatory kernel has been drawn with only two E excitatory neurons, but in reality they represent a population of perhaps 40 or 50 interneurons. These interneurons are glutamatergic and activate both NMDA and AMPA receptors. There are a number of complementary processes that make the neurons cease being active at the end of a burst, and conversely to recruit neurons at the onset of a burst.

In a network generating alternating activity—for instance, on two sides of a segment—there are two burst generating kernels, one on each side, connected with inhibitory neurons (Fig. B27.1). Thus when one kernel of excitatory neurons is active, the contralateral kernel is inhibited, and when its burst is terminated, the other side takes over due to the background drive from the brainstem. The crossed inhibitory neurons thus are responsible for the coordination beween the right and left sides, and they use glycine as a transmitter.

A generalizing conclusion is that in a number of burst generating circuits for, for example, different forms of locomotion, respiration, and saccadic eye movements, burst generating kernels of excitatory interneurons play a key role. In all cases the synaptic interaction is critical, and the cellular properties contribute significantly.

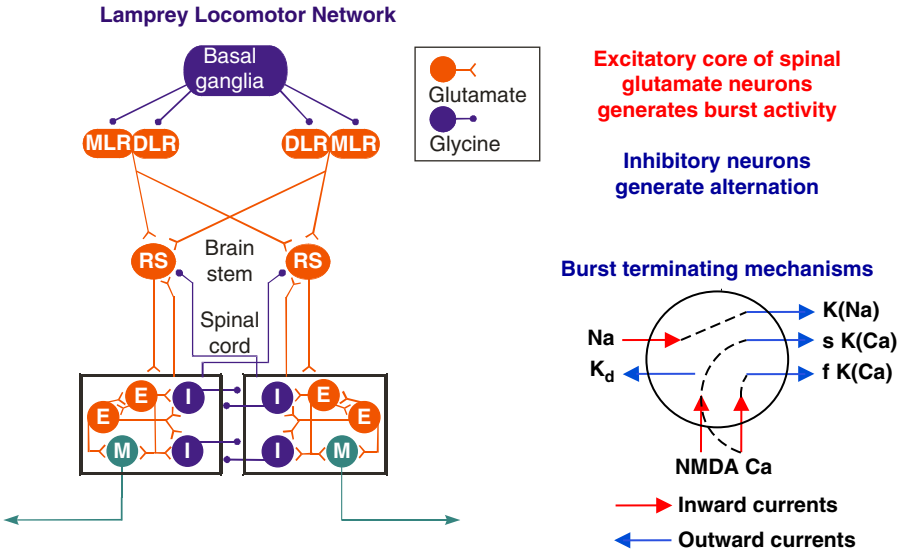


FIGURE B27.1 **The segmental locomotor network of the lamprey.** Schematic representation of the different components of the core neural circuitry that generates rhythmic locomotor activity. All neuron symbols denote populations rather than single cells. The reticulospinal (RS) glutamatergic neurons excite all types of neurons at the segmental level. The excitatory (E) neurons excite all three groups of neurons indicated, and the inhibitory neurons inhibit contralateral neurons and motoneurons (M). The reticulospinal neurons in turn receive excitatory drive from the mesencephalic (MLR) and diencephalic (DLR) locomotor regions, which at rest are under tonic inhibitory control from the output cells of the basal ganglia. When locomotor activity is to be initiated, the inhibitory control from the basal ganglia is removed (see above), and the network is free to operate. To the right, cellular mechanisms are indicated that aid in burst termination. During each burst, Ca^{2+} and Na^+ ions accumulate in the cell that in turn activate calcium and also sodium-dependent K^+ channels that progressively hyperpolarize the cells, thereby terminating ongoing spike activity.

Rhythmic Movements

1. Walking movements (or locomotor behaviors in general) are produced by CPGs in the spinal cord (or ganglia of invertebrates). These spinal CPGs are turned on by descending control signals from particular areas in the brainstem (locomotor areas), which also determine the level of locomotor activity (e.g., slow walking versus trot or gallop). The CPG then sequentially activates motoneurons/muscles that support the limb during the support phase and then move it forward during the swing phase. Sensory stimuli from the moving limb also contribute importantly to regulate the duration of the support phase and the degree of activation of different muscles that are sequentially activated in each step cycle. The relative importance of the sensory component varies between species and with the speed of locomotion. During very fast movements, there is actually no time for sensory feedback to act, and adjustments need instead to be performed in a predictive mode.
2. Chewing movements are controlled by brainstem circuits that in principle are designed in a similar manner to those of locomotion and generate alternating activity between jaw-opener and -closer muscles.
3. Breathing movements are continuously active from the instant of birth to the time of death, except for short voluntary interventions when speaking, singing, or choosing not to breathe for some other reason, like diving. The level of respiration (depth and frequency) is driven by metabolic demands (e.g., pCO_2) detected by chemosensors in the brainstem, where the respiratory CPG is also located. Sensory stimuli from lung volume receptors contribute by setting the level at which the inspiration is terminated and expiration takes over.

Eye Movements

Fast saccadic eye movements are used when looking rapidly from one object to another and are represented with a much more complex organization in the superior colliculus (mesencephalon) than the motor patterns discussed earlier. Saccadic eye movements with different directions and amplitudes can be generated by the stimulation of different microregions in the superior colliculus, according to a well-defined topographical map. Each microregion is responsible for activating a subset of brainstem interneurons, which in turn activates the appropriate combination of eye motoneurons in order to move the two eyes in a coordinated fashion from one position to another. The purpose is to bring the visual object of interest into the foveal region so that it can be scrutinized in the greatest detail. To recruit a saccadic eye movement to a specific site, cortical or subcortical areas have to access the appropriate microregions within the superior colliculus to trigger the specific eye movement desired.

Other types of eye movement control are used when tracking an object that is moving, as when watching a game of tennis. These tracking movements are represented in the cerebral cortex and the brainstem.

Posture

Animals, including humans, have the ability to position the body in a variety of postures. The most stable position is lying horizontally in bed, which requires little activity from the nervous system. The situation is very different when standing and even more so when supporting oneself on one leg. This requires a very structured command to handle the activation of different muscles in the legs, trunk, shoulders, and neck. The nervous system issues this complex motor command to achieve a certain position. The ability to maintain a stable position standing on one leg is then critically dependent on feedback from a variety of sensory receptors that detect falling to one side or the other. These receptors activate reflex circuits that try to counteract the impending loss of balance. Of great importance are the vestibular receptors located in the head, which sense head movements in the different directions. Vestibular reflexes act on the neck muscles to correct the position of the head, and vestibulospinal reflexes act on the trunk and the extensor muscles of the limbs. In addition, vestibular effects are also mediated indirectly by the reticulospinal system. Muscle receptors in neck and limb muscles are of equal importance. They also play a critical role in detecting and compensating for postural perturbations, in particular around the ankle, where skin receptors on the bottoms of the feet are activated. Finally, visual feedback may contribute as when one is standing on a moving surface such as the deck of a boat.

One can decide at will what posture to assume, whether standing on one leg or sitting in an armchair. In that sense, assuming different postures represents a set of voluntary motor acts organized at both the brainstem and forebrain level. At the same time, any given posture is stabilized in an "automatic" way by the sensory mechanisms described above.

ROLES OF DIFFERENT PARTS OF THE NERVOUS SYSTEM IN THE CONTROL OF MOVEMENT

The different basic motor patterns are present in different forms in most vertebrates and invertebrates. In the former, the underlying neural networks are located either in the brainstem or in the spinal cord, and in the

latter, in the chain of ganglia. In vertebrates, from fish to mammals the basic organization of the nervous system is similar with regard to the spinal cord, brainstem (medulla oblongata and mesencephalon), and diencephalon. Each species and larger group of vertebrates may have specializations related to particular types of functions, and complexity increases during evolution. The cerebral cortex, with its distinct lamination, is present in mammals and most highly developed in primates, including humans (Fig. 27.5).

Basic Motor Programs Are Located at the Brainstem–Spinal Cord Level

The brainstem–spinal cord is, to a large extent, responsible for the coordination of different basic motor patterns. The spinal cord contains the motor programs (CPGs) for protective reflexes and locomotion, whereas those for swallowing, chewing, breathing, and fast saccadic eye movements are located in the brainstem (mesencephalon and medulla oblongata). In most cases, however, both the brainstem and the spinal cord are involved to some degree. In mammals as well as lower vertebrates, the brainstem–spinal cord, isolated from the forebrain (di- and telencephalon), is able to produce breathing movements and swallowing, as well as walking and standing. These brainstem–spinal cord animals (referred to as decerebrate models) can thus be made to walk, trot, and gallop with respiration adapted to the intensity of the movements and to swallow when food is put in their mouth. However, they perform these maneuvers in a stereotyped fashion, such as a robot or a reflex machine. The movements are thus not goal directed nor adapted to the surrounding environment, but are nevertheless coordinated in an appropriate way.

Diencephalon and Subcortical Areas of Telencephalon Are Involved in Goal-Directed Behavior

Mammals that are devoid of the cerebral cortex but have the remaining parts of the nervous system intact have been used to investigate the importance of these areas. Such animals display a more advanced behavioral repertoire when compared to the brainstem animals described earlier. At first glance, these decorticate animals look surprisingly normal. They move around spontaneously in a big room and can avoid obstacles to some extent. They eat and drink spontaneously and can learn where to obtain food and search for food when supposedly hungry. They may also display emotions such as rage and attack other animals; however, they appear unable to interact in a normal way with other individuals of the same species. Even without the cerebral cortex, a surprisingly large part of the normal motor repertoire can be performed, including some aspects of goal-directed behavior. Thus the cerebral cortex is not required to achieve this level of complex motor behavior (however, see later discussion). The diencephalon and subcortical areas of the telencephalon of the forebrain contain two major structures that are important in this context: the hypothalamus and the basal ganglia.

Hypothalamus

The hypothalamus is composed of a number of nuclei that control different autonomic functions, including temperature regulation and intake of fluid and food. The latter nuclei become activated when the osmolality is increased (fluid is needed) or the glucose levels become low (food is required). Stimulation of the paraventricular nucleus involved in the control of fluid intake results in a sequence of motor acts involving a number of different motor programs. Continuous activation of this nucleus by electrical stimulation or local ejection of a hyperosmolalic physiological solution leads to an alerting reaction. The animal first starts looking for water (perhaps experiencing a feeling of thirst). It then starts walking toward the water, positions itself at the water basin, bends forward, and starts drinking. The animal will continue drinking as long as the nucleus is stimulated. These hypothalamic structures are thus able to recruit a sequence of motor acts that appear in a logical order. Other parts of this region can elicit rage and attack behavior, as was first described by the Nobel laureate Walter Hess. Humans have seven distinct motor programs to express different types of emotions from crying (already present directly after birth) to smiling, laughing and the display of anger. These motor programs are inborn but what you laugh at is of course culturally bound.

Basal Ganglia

Basal ganglia constitute a second major set of nuclei that is conserved from lamprey to man. They are of critical importance for the normal initiation of motor behavior. They are subdivided into an input region (the dorsal and ventral striatum) activated by the cerebral cortex, thalamus, and an output region referred to as pallidum (globus pallidus interna and substantia nigra reticulata). The output neurons are inhibitory and have a very high level of activity at rest. The purpose is to keep the different brainstem motor centers under tonic inhibition so that they are not activated under resting conditions and also thalamic neurons projecting back to the motor areas of the cerebral cortex. The input stage of the striatum determines the level of activity in the different pallidal output neurons. When a motor pattern, such as a saccadic eye movement, is going to be initiated, the pallidal output

neurons that are involved in eye motor control become inhibited by the striatum. This means that the tonic inhibition produced by these neurons at rest is removed, and the saccadic motor centers in mesencephalon (superior colliculus) are relieved from tonic inhibition and become free to operate and induce an eye saccade to a new visual target.

In contrast to the tonically active output neurons of the basal ganglia (pallidum), the striatal neurons are silent at rest and have membrane properties that make it difficult to activate them from cortex or thalamus, unless the dopamine system is active. The level of activity in dopaminergic neurons in the substantia nigra and associated nuclei determines the responsiveness of striatal neurons. If the level of dopamine activity is reduced, as in Parkinson's disease, it becomes very difficult to activate the striatal neurons and thereby to initiate and carry out most types of movements. This results in a severe handicap characterized by paucity of movement (hypokinesia). The converse condition with enhanced levels of dopamine in the basal ganglia (which can occur as a side effect of medication) results instead in a richness of movements, and even in unintended movements (hyperkinesias). These movements may be well coordinated, but without a purpose, and occur without the conscious involvement of the patient. Dopamine neurons are thus of critical importance for the operation of the striatum and they regulate the responsiveness of the striatal circuitry to input from the cerebral cortex and thalamus. Striatum in itself clearly has a key role for the operation of the entire motor system.

The Cerebral Cortex and Descending Motor Control

In the frontal lobe there are several major regions that are involved directly or indirectly in the execution of different complex motor tasks like Broca's area for speech, or the skilled movements used to control hands and fingers when writing, drawing, or playing an instrument. Some neurons in the frontal lobe are activated not only when we perform a given movement, but also when we see somebody perform the same type of movement. They are therefore called mirror-neurons and may underlie the ability to imitate. Many of these regions are organized in a somatotopic fashion. In the largest area, referred to as the primary motor cortex or M1 in the precentral gyrus, the legs and feet are represented most medially and the trunk, arm, neck, and head are represented progressively more laterally. Areas taken up by the hands and the oral cavity are very large in humans, and are much larger than that for the trunk. This is explained by the fact that speech and hand motor control require a greater precision and thus a larger cortical processing area than the trunk. The latter is important for postural control but is less involved in the type of skilled movement controlled by the motor cortex (Fig. 27.6).

The large pyramidal cells in the motor cortex send their axons to the contralateral side of the spinal cord and are able to activate their target motoneurons directly, but a number of interneurons in the brainstem or spinal cord are also influenced. These long projection neurons are called corticospinal neurons. Corticospinal neurons of the arm region in M1 thus project to arm motoneurons in the spinal cord, as well as interneurons involved in the control of arm, hand, and fingers. In addition to M1, other areas in the frontal lobe, such as the supplementary motor area and prefrontal areas, are involved in other aspects of motor coordination (Fig. 27.7).

The cortical influence on movements is exerted through several different channels, via direct projections to the input stage of the basal ganglia (striatum), through effects exerted via brainstem circuits and as mentioned above also directly to the spinal cord via corticospinal projections. The brainstem contains a number of descending pathways (rubrospinal, vestibulospinal, and reticulospinal pathways) that can initiate movements and correct

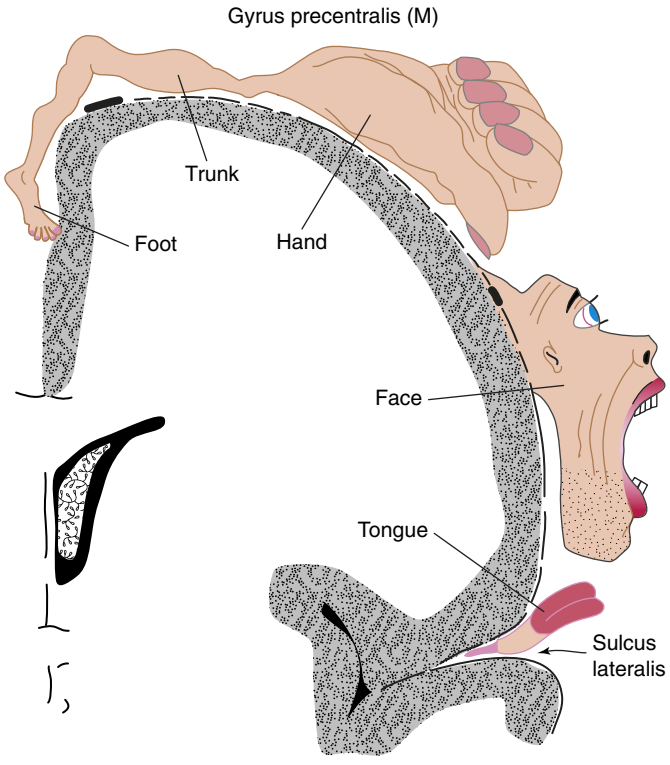


FIGURE 27.6 Organization of the primary motor cortex (M1). The different parts of the body are represented in a somatotopic fashion in M1, with the legs represented most medially, arms and hand more laterally, and the oral cavity and face even more so. Note that the two areas that are represented with disproportionally large regions are the hand with fingers and the oral cavity. They are represented in this way due to the fine control required in speech and fine manipulation of objects with the fingers. *Adapted from Penfield and Rasmussen (1950).*

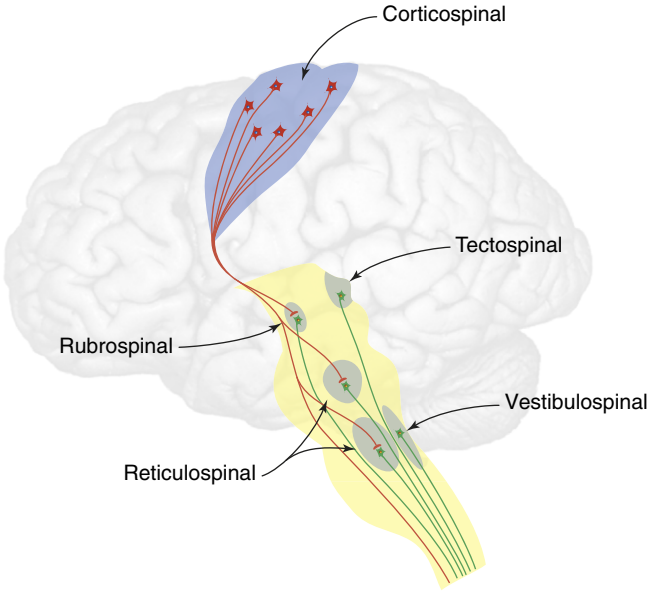


FIGURE 27.7 Descending pathways projecting from the brain that mediate motor actions to the spinal cord. From the cerebral cortex, including M1, a number of neurons project directly to the spinal cord to both motoneurons and interneurons. In addition, they also project to different motor centers in the brainstem, which in turn sends direct projections to the spinal cord. One of these centers is the red (rubrospinal) nucleus, which influences the spinal cord via the rubrospinal pathway. Corticospinal and rubrospinal pathways act on the contralateral side of the spinal cord and have somewhat overlapping functions. They are sometimes referred to as the lateral system, as most of the fibers descend in the lateral funiculus of the spinal cord. Vestibulospinal and reticulospinal pathways are important both for regulating posture and for correcting perturbations, and they mediate commands initiating locomotion. The different reticulo- and vestibulospinal pathways are sometimes lumped together as the medial system, as most of the axons project in the medial funiculus of the spinal cord.

motor performance, while others provide more subtle modulation of the spinal circuitry. The former act rapidly in the millisecond time frame that is required in motor control. Examples of the latter are the slow-conducting and slow-acting noradrenergic and serotonergic pathways. These pathways set the responsiveness of different types of neurons, synapses, and spinal circuits. The cortical control of motor coordination is, in addition to the effects via the basal ganglia, achieved through both direct action on spinal and brainstem motor centers but also to a significant degree by parallel action on a variety of brainstem nuclei. The execution of movements that are initiated from the cortex is, to a large degree, a collaborative effort of many parts of the nervous system. The reticulospinal and vestibulospinal pathways mediate the control of posture, and the former take part in the initiation of locomotion, via the CPG located in the spinal cord (Fig. 27.8).

Judging from experiments on mammals and primates and on patients that have suffered focal lesions of the frontal lobe, the cortical control of movement is of particular importance for dexterous and flexible motor coordination, such as the fine manipulatory skills of fingers and hands and also for speech. It is very difficult for a casual observer to see the difference between a normal monkey moving around in a natural habitat and a monkey that is lacking the corticospinal direct projections to the spinal cord but has all other cortical circuits intact. They are able to move around, climb, and locomote as normal monkeys. It is only when special tests are performed that one can see that delicate independent movements of the individual fingers are incapacitated. This is a type of motor coordination that is needed in monkeys and other primates for skilled manipulations of the environment, such as picking fine food objects from small holes.

The most flexible types of motor coordination, such as the skilled movements of the fingers and hands or in speech, involve the frontal lobes. These types of movements are often referred to as voluntary because they are performed at will. This terminology is commonly used, although inappropriate in that more basic types of motor coordination, such as the ones used when walking, chewing, or positioning ourselves in a relaxed posture, are also controlled by will, although motor coordination is handled to a larger degree by motor programs located at the brainstem–spinal cord level. Thus it appears important not to draw a clear-cut dividing line between the different types of movements, at least with respect to the voluntary aspect of their control.

It is also useful to realize that, to some degree, simple reflexes can also be modulated by will. For instance, the simple skin avoidance reflex (flexion reflex) that causes rapid withdrawal of a finger after touching something painful is subject to such control. Knowledge that the object will be sufficiently hot to be somewhat painful enables this modulation if it is important. If, however, the object is touched without knowing that it is hot, the hand will be withdrawn with the shortest possible latency. Thus, in the entire motor apparatus from the simplest reflex to a skilled movement, there is the possibility for modulation and flexibility. The human nervous system allows a remarkable combination of movements, adapted to different situations. This flexibility is perhaps greater for primates and humans than for any other species. However, other species may perform much better in particular types of specialized movements. A cheetah runs faster, a hawk catches prey at higher speeds, and a fly walks upside down on the ceiling.

Visuomotor Coordination—Reaction Time

Visuomotor coordination is amazing. Consider an approaching object like a ball that you would like to catch. This requires a rapid calculation of the path of the ball and being able to extend the arms and position the hands in the appropriate spot at just the right time. The motor command needs to consider not only the timing but also the

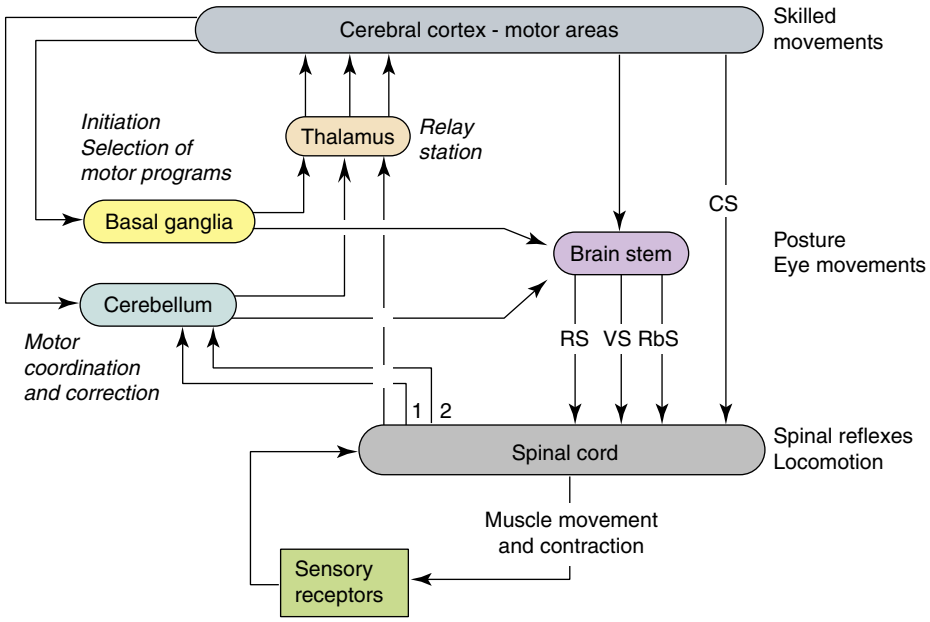


FIGURE 27.8 Summarizing scheme of the interaction between different motor centers. The different major compartments of the motor system and their main pathways for interaction are indicated. The basic functions of the different compartments and descending tracks are summarized on and below the scheme. CS, corticospinal; RbS, rubrospinal; VS, Vestibulospinal; RS, reticulospinal.

Basic functions of descending tracts

Cortico- and rubrospinal
1. Transmission of commands for skilled movements.
2. Corrections of motor patterns generated by the spinal cord.

Reticulospinal
1. Activation of spinal motor programs for stepping and other stereotypic movements.
2. Control of upright body posture.

Vestibulospinal
Generation of tonic activity in antigravity muscles

visuomotor details about joint angles from the eyes to the hand via the neck and head. How is this achieved? The parietal lobe processes visual information that will contribute to the initiation of the motor commands.

In order to arrive at an accurate and precise motor command, a wealth of information needs to be processed about the dynamic and static conditions of the different parts of the body in relation to each other and to the surrounding world. The processing time has been investigated experimentally by having a subject respond to a signal by performing a movement, such as pressing a button when a light goes on. The delay is around 0.1–0.2 s and is called the simple reaction time. It represents a minimal delay for a given test situation that cannot be shortened by training (Fig. 27.9).

The more complex the situation is, the longer the reaction time. If a choice is involved, such as responding by pressing different buttons to light stimuli of different colors, the time delays become much longer than in the simple reaction time task. The choice reaction task time increases in proportion to the level of complexity and the number of choices (Fig. 27.9).

When young children are tested, their reaction times are much longer than in adults. The simple reaction time for a 6-year-old can be three times that of a 14-year-old, and the difference with choice reaction times may be even larger. In everyday life this condition may have severe consequences, such as in an unexpected traffic situation. A 6-year-old will need at least three times more time to interpret what he or she sees. This is the basis for recommending that children below 11 years of age should not ride a bicycle in open traffic. It is noteworthy that this is a gradual maturation process, and training will not shorten the reaction time. During aging the choice reaction time tends to increase again, a factor that can be important when driving a car.

The cerebellum is also involved in the coordination of movement. Although it is smaller than the cerebral cortex

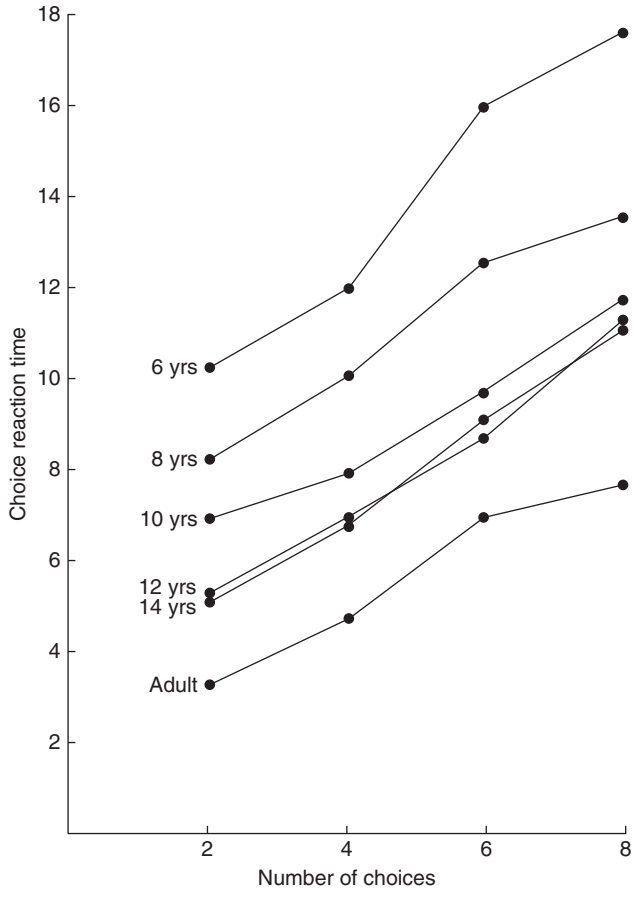


FIGURE 27.9 Reaction time tasks of different complexity and in different age groups. The choice reaction time is plotted versus the number of choices the subject is exposed to for six different age groups. A simple experimental situation with two choices takes an adult 0.3 s to initiate, but a 6-year-old needs around 1.0 s. A 10-year-old falls in between at around 0.7 s. An increase in the number of choices causes a marked increase in reaction time. From a practical standpoint, this provides important information when one considers the possibility for a child to cope with different demanding situations, such as moving around in a modern city. *Data replotted from Conolly (1970).*

in volume, it is thought to contain as many nerve cells as the latter. It has a very stereotyped neuronal organization with two types of inputs from mossy and from climbing fibers. These inputs not only carry information from the spinal cord about ongoing movements in all different parts of the body, but they also carry information from the different motor centers about planned movements even before a movement has been executed. The cerebellum also interacts with practically all parts of the cerebral cortex. This means that it is updated continuously about what goes on in all parts of the body with regard to movement and also about the movements that are planned in the immediate future.

Lesions of the cerebellum lead initially to great problems with postural stability and with a lack of accuracy of movements. Most movements can be carried out, but their quality is reduced drastically. The cerebellum was originally thought to be involved exclusively in the coordination of movement. In addition to motor control, evidence has accumulated that the lateral parts of cerebellum may also be involved in different cognitive tasks.

The cerebellum is subdivided into a great number of different microregions that each process different kinds of information. These different regions respond via their output neurons, called Purkinje cells. The cerebellar cortex contains only inhibitory interneurons, and the two input systems, climbing fibers and mossy fibers (via granular cells), provide the excitatory components. All Purkinje cells are inhibitory and project to the cerebellar nuclei, which in turn are excitatory and project to different motor centers in the brainstem and to the cortex via the thalamus.

Each granular cell receives very specific input from only a few mossy fibers, which are able to make them fire. The granular cell axons form parallel fibers in the cerebellar cortex, and thousands of parallel fibers impinge on each Purkinje cell that has a very extensive dendritic tree. It is believed that only a small portion of these parallel synapses are functional in any given moment but that they are available to become recruited in learning situations. In contrast, there is only one climbing fiber for each Purkinje cell. The climbing fibers can serve as "error detectors," at least under some conditions. Through the effects of climbing fibers on the Purkinje cells, the synaptic efficacy of the parallel fiber input to the Purkinje cell can be modified (depressed). This change in efficacy can last over many hours and possibly much longer and is referred to as long-term depression (LTD). It is generally thought that this LTD can contribute to motor learning (see later). With regard to basic movements such as posture, walking, and the much-studied eye blink reflex, the cerebellum is clearly involved through each movement phase. It is likely that the fine tuning and modification of movements that are required in different behavioral contexts involve the cerebellum. With regard to the eye blink reflex, motor learning has been demonstrated in terms of associating an unconditioned stimulus with a conditioned one. It requires that the cerebellar cortex be intact, which strongly suggests that the cerebellar cortex contributes importantly to this type of motor learning.

There is still, however, much to learn about the actual processing that goes on in the cerebellum, a structure that regulates the quality of motor performance and is involved in some forms of motor learning. A variety of complex motor tasks, including speech and handwriting, can still be performed, but with less accuracy. Thus, the motor programs for these learned tasks cannot be stored exclusively in the cerebellum. When new types of movements are learned, such as riding a bicycle, playing an instrument, or typing at a keyboard, the new stored programs for sequences of motor acts must also involve other

structures in the brain, most likely the basal ganglia and the cerebral cortex.

Motor Learning

All vertebrate and invertebrate species with a nervous system have a motor infrastructure that is determined evolutionarily. In the case of primates, including humans, there are preformed circuits (e.g., CPGs) that allow performance of a basic movement repertoire (e.g., locomotion, posture, breathing, eye movements), as well as basic circuits that underlie reaching, hand and finger movements, and sound production, as in speech. This constitutes the motor infrastructure that is available to a given individual after maturation of the nervous system has occurred.

The motor system is in addition characterized by its ability to learn, in what is referred to as motor or procedural learning. The cortex and basal ganglia play an important role for learning of new motor sequences and skills, and the cerebellum is indispensable for the quality of the different motor patterns and the ability to associate two stimuli, as during conditioned reflexes. When one is able to perform a movement with great precision, like hitting a ball, this may lead to an activation of the dopamine system and a feeling of satisfaction (reward). It has a marked ability to facilitate the development of synaptic plasticity, which is learning. It facilitates the particular synapses that were active as the successful movement was performed.

One can also learn motor sequences, as when playing the flute. The particular finger settings that produce a given tone can be retained in memory, along with the sequence of tones that produce a certain melody. Thus particular muscle combinations that produce a sequence of well-timed motor patterns are stored. Similarly, when learning to pronounce different sounds (phonemes) and combine them into words, the muscle activation patterns are stored. A simpler case occurs when learning to recombine basic movement patterns to be able to maintain equilibrium while bicycling. There are a large number of examples involving different types of motor learning and parts of the body.

It is characteristic of a given learned motor program that one can "call" upon it to perform a given motor act over and over again. Despite this, there is generally little awareness of the way in which the movement is actually performed in terms of the muscles involved and their timing. Information has thus been stored in the nervous system with regard to which parts of the infrastructure of the motor system should be used to produce a given learned motor pattern, as well as the exact timing between the different components. However, the details of the complex storage process are not yet fully understood.

CONCLUSION

After this overview of the motor control system, it should be apparent that biological evolution has succeeded in refining remarkable sensorimotor machinery capable of executing the motor tasks required for the full behavior of a given species. The degree of difficulty involved is apparent when watching the clumsiness and lack of flexibility in robots, characteristic of even the most advanced models that have been developed. Details of the different neural subsystems involved in motor control are dealt with in the following chapters.

Suggested Readings

Dean, P., Porrill, J., Ekerot, C. F., & Jörntell, H. (2010). The cerebellar microcircuit as an adaptive filter: experimental and computational evidence. *Nature Reviews Neuroscience, 11*(1), 30–43.

Deliagina, T. G., Orlovsky, G. N., Zelenin, P. V., & Beloozerova, I. N. (2006). Neural bases of postural control. *Journal of Physiology (Bethesda), 21*, 216–225.

Georgopoulos, A. P. (1999). News in motor cortical physiology. *News in Physiological Sciences, 14*, 64–68.

Grillner, S. (2006). Biological pattern generation: The cellular and computational logic of networks in motion. *Neuron, 52*, 751–766.

Grillner, S., Hellgren, J., Menard, A., Saitoh, K., & Wikstrom, M. A. (2005). Mechanisms for selection of basic motor programs—roles for the striatum and pallidum. *Trends in Neurosciences, 28*, 364–370.

Ito, M. (2002). The molecular organization of cerebellar long-term depression. *Nature Reviews Neuroscience, 3*, 896–902.

Johansson, R. S., & Flanagan, J. R. (2009). Coding and use of tactile signals from the fingertips in object manipulation tasks. *Nature Reviews Neuroscience, 10*(5), 345–359.

Marder, E. (1998). From biophysics to models of network function. *Annual Review of Neuroscience, 21*, 25–45.

Redgrave, P., Rodriguez, M., Smith, Y., Rodriguez-Oroz, M. C., Lehericy, S., Bergman, H., et al. (2010). Goal-directed and habitual control in the basal ganglia: Implications for Parkinson's disease. *Nature Reviews Neuroscience, 11*(11), 760–772.

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. *Nature Reviews Neuroscience, 11*, 264–274.

Rossignol, S., Dubuc, R., & Gossard, J. P. (2006). Dynamic sensorimotor interactions in locomotion. *Physiological Reviews, 86*(1), 89–154.

Shepherd, G. M., & Grillner, S. (2010). *Handbook of brain microcircuits*. New York: Oxford University Press.

Stephenson-Jones, M., Samuelsson, E., Ericsson, J., Robertson, B., & Grillner, S. (2011). Evolutionary conservation of the basal ganglia as a common vertebrate mechanism for action selection. *Current Biology, 21*, 1081–1091.

Sten Grillner

This page intentionally left blank

CHAPTER

28

The Spinal and Peripheral Motor System

Most coordinated movements involve the use of several muscles, each with a precisely orchestrated timing and pattern of activation and relaxation. Contraction of an individual muscle is controlled by a pool of motor neurons that innervate that muscle. Patterned contraction of several muscles requires coordination between pools of motor neurons. For many movements, interneurons in the spinal cord provide the means for coordinating activity of different motor neuron pools. A key feature of spinal interneurons is their convergent and divergent connections that allow them to integrate signals from different sources. Interneurons that receive inputs from descending motor systems and sensory afferents, for example, can update a voluntary command for limb movement to incorporate feedback from the moving limb. Spinal interneurons also form connections with other spinal interneurons to create networks that generate rhythmic patterns of excitatory and inhibitory activity. These networks play a key role for organizing rhythmic, repetitive movements such as scratching or locomotion. In this chapter, we will use locomotion as an example of a motor behavior to illustrate how the neural elements of the spinal cord and peripheral neuromuscular system operate together to carry out a movement. As details of the motor neurons, spinal interneurons, muscles, and sensory afferents are presented, the primary focus will be on mammals, but variations that occur in other vertebrates and in invertebrates are highlighted in Box 28.1.

LOCOMOTION IS A CYCLE

Locomotion is the act of moving from place to place. Walking, swimming, and flying are modes of locomotion. All modes of locomotion consist of a cycle of patterned and rhythmic activity of different muscle groups. Locomotion can be described by the phase relationships between limbs and between muscle groups during the cycle. Nonlimbed vertebrates such as lampreys swim by alternately contracting and relaxing trunk muscles on the two sides of the body; the sides are in opposite phases of the cycle. Along the length of the lamprey's body, there is a slight phase lag between segments, creating the propulsive wave that moves it forward. In contrast, birds activate their wings synchronously in flight; the limbs are in phase, with no lag.

Stepping, the major form of mammalian locomotion, is a cycle with two phases: stance (the extension phase) and swing (the flexion phase). In people, the stance phase begins when the heel strikes the ground and the leg begins to bear weight. As stance progresses, the weight is gradually shifted forward, rolling from the heel to the ball of the foot. The swing phase begins when the toe leaves the ground and the leg moves forward. To accomplish this motor sequence within one leg, muscles acting across the hip, knee, or ankle are activated in a characteristic pattern (Fig. 28.1). In general, muscles that flex the joint are more active during the swing phase and extensor muscles during the stance phase. However, within each phase, the individual muscles have a more elaborate and finely graded timing and pattern of activation. Stepping can also be performed with different gaits. Gaits can be characterized by the cycle rate, the relative proportion of the cycle spent in swing and stance phases, and the phase relationships between limbs. At slow or moderate walking speeds (i.e., slow cycle rates), the extensor phase is longer than flexion and the extension phases of opposite legs overlap. In a walking gait both feet are in simultaneous contact with the ground for some period during the cycle. As the speed of stepping increases and the cycle is shorter, the duration of the extension phase shortens, with briefer periods of double foot contact. At high speeds where there is no overlap of extension phases between the legs, the gait switches recognizably from walking to running.

© 2013 Elsevier Inc. All rights reserved.

BOX 28.1

SPECIES DIVERSITY IN NEUROMUSCULAR SYSTEMS

Not all animals control muscle contraction in the same way as mammals. Differences are found in the proteins and structure of the muscle fiber, the number of neuromuscular junctions on each muscle fiber, the neurotransmitters used at the neuromuscular junction, and the number of motor neurons contacting each muscle fiber. Major differences are found in invertebrates, which often have evolved adaptations that suit their movement patterns or constraints imposed by their environments.

Some invertebrates require muscles that generate great forces or provide mechanical stiffness. One solution used by mussels, the nematode *C. elegans,* and the fruit fly *Drosophila* is the incorporation of molecules of paramyosin into thick filaments, so that additional cross-bridges can be formed. Mussel retractor muscles also have a specialized catch property, in which cross-bridges are tightly bound with low energy expenditure until the catch state is released by neurotransmitter action that leads to phosphorylation of twitchin, a titin-like molecule. Very slow movements can be achieved by long sarcomeres, since Ca^{2+} diffusion throughout the sarcomere takes longer. In mammalian muscle fibers, sarcomeres are generally 2–3 microns long, although longer sarcomeres are occasionally found in muscles that stiffen but do not twitch, such as the tensor tympani muscle of the cat eardrum. In contrast, claw muscles of crabs may have sarcomeres from 6 to 18 microns, and one marine worm was found to have sarcomere lengths of 30 microns.

Most mammalian muscles usually contain a mixture of muscle fiber types, but some animals have distinct muscles for fast and slow movements. Trout, for example, use red muscles composed of slow twitch fibers for sustained swimming and white, fast twitch muscles for leaping and rapid bursts of speed.

Mammals and birds use acetylcholine as the transmitter at the neuromuscular junction. So does *C. elegans*, but many other invertebrates use glutamate, including arthropods such as *Drosophila*, locust, and crayfish. Co-release of glutamate and peptides occurs at some neuromuscular junctions, such as the peptide FMRF-amide in the sea snail *Aplysia*. Crustacean muscles have separate excitatory and inhibitory neuromuscular junctions that use glutamate and GABA.

In mammals, most muscle fibers are innervated by a single motor neuron, with a few exceptions, such as laryngeal muscles and spindles. Innervation by more than one motor neuron is seen in some avian muscles, and many invertebrate muscles. The motor neurons may merely have an additive function, as in *Drosophila,* where recruitment of motor neurons produces graded contractions of the muscle. In other animals, the different motor neurons may play distinct roles. The locust jumping muscle has a dual arrangement, with one motor neuron that produces a rapid and powerful excitation and another that produces graded excitation and contraction.

Lastly, the central control of motor neurons differs. In mammals, a motor neuron action potential leads inevitably to a muscle fiber action potential. In crayfish, presynaptic inhibition of motor terminals allows modulation at a more distant site. Structural specializations, such as electrical synapses onto motor neurons and gap junctions between motor neurons, often occur in invertebrates, but generally are found only during development in mammals. The distributed structure of invertebrate systems and rich palette of ion channels expressed by motor neurons is favorable to the formation of motor pattern generating circuits.

Mary Kay Floeter

Central Pattern Generating Circuits

Although we are able to exert voluntary control over locomotion, the basic patterns of movement for locomotion can be generated by the spinal cord itself, without descending (from the brain) or sensory (from the periphery) inputs. The first demonstration came from Graham Brown (1911), who observed that cats in which the spinal cord and dorsal roots were cut were still able to walk on a treadmill as long as their weight was supported. Because there was no sensory feedback or descending input, he suggested that the rhythmic activity in the hindlimb muscles must be generated by a motor program within the spinal cord. The motor program, also known as a central pattern generator (CPG), needed to explain three important features of locomotion: first, its rhythmic and cyclical occurrence; second, coordinated alternation between flexor and extensor activity within the same limb; and third, alternation between opposite limbs. Brown proposed that the CPG consisted of two mutually inhibitory "half-centers," both under tonic excitation: one half-center drove muscles that flexed the legs and the other drove muscles that extended the legs. The mutual inhibition between the two half-centers would cause activity to oscillate between them, producing alternating limb flexion and extension.

CPGs are now thought to consist of networks of neurons that are synaptically or electrically connected,

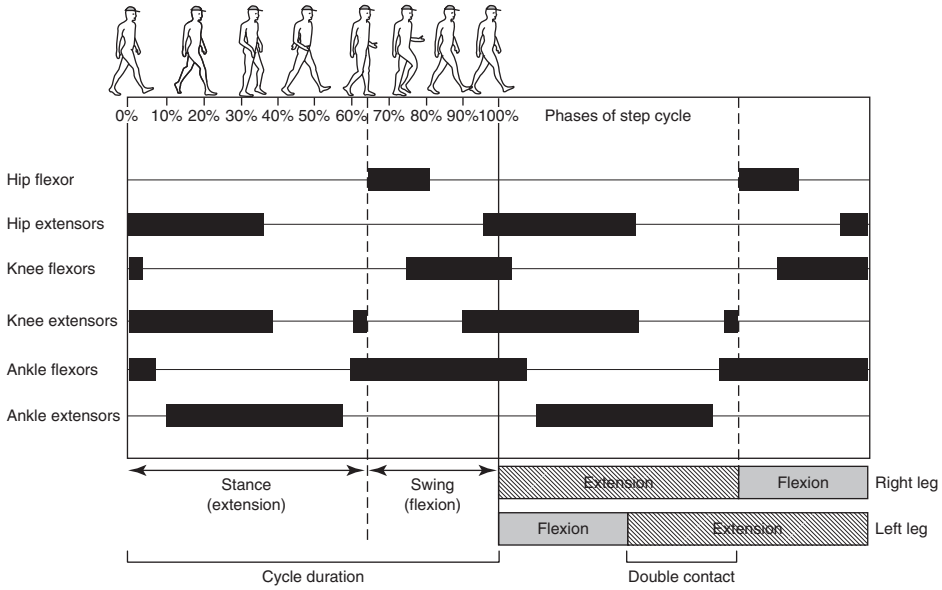


FIGURE 28.1 Schematic diagram of two step cycles in human locomotion. Drawings at the top illustrate the position of the limbs during one cycle. Dark horizontal bars below the drawings show the timing of activity in flexor and extensor muscles of the right leg. The division of each step cycle into stance (extension) and swing (flexion) phases is indicated at the bottom of the figure. During walking, the stance phases of the two legs overlap, with both feet in contact with the ground.

with cellular properties that generate rhythmic firing. CPGs are found in many areas of the central nervous system and are responsible for different types of patterned rhythmic activities such as breathing, chewing, swallowing and walking. The identification and characterization of the neurons that comprise the locomotor CPG in the spinal cord is an active area of research that will be discussed later in this chapter.

Sensory Contributions to Movement

During a movement, the sensory afferents from muscles, joints, and skin will each be activated in a distinct way, reflecting how that movement displaces joints, produces pressure on regions of skin, and stretches individual muscles. For rhythmic movements such as walking, these recurring "multisensory" feedback signals often reinforce activity of CPG circuits. However, when a perturbation occurs during the movement, the resulting sensory signal may evoke a reflex motor response that alters the movement. Many of these reflex responses adapt the movement in an appropriate way to proceed under the changed environmental conditions. Most sensory afferents produce these reflex effects through interneuronal circuits. (The stretch reflex is an exception.) The excitability of many of these interneuronal circuits is usually modulated during movement, allowing reflexes to be suppressed or enhanced at different phases. CPGs are one source of this phase modulation. Many reflexes can also be elicited when the organism is not moving, and in these situations the reflex tends to exhibit a more stereotyped motor pattern.

CONNECTING THE SPINAL CORD TO THE PERIPHERY

Ultimately, movement is carried out by muscles, and the motor commands from the central nervous system must be transmitted to motor neurons that innervate the muscles. The organization of motor neurons into pools, and the manner in which motor neurons' properties are matched to contractile and metabolic properties of the muscle fibers simplify many of the variables needed to be controlled to produce smooth movements with precise amounts of force. Feedback from sensory receptors of the muscle play a key role in monitoring the force and stretch of the muscle. This section will cover some of the basic properties of motor neurons, muscle, and muscle receptors.

Motor Neurons

Motor neurons are located in the ventral horn of the spinal cord. The pool of motor neurons that innervates an individual muscle forms an elongated column that typically extends over two or three spinal cord segments. Intermingled among the large α-motor neurons that innervate skeletal muscle fibers are smaller γ-motor neurons, which innervate muscle spindles, as well as β-motor neurons, which dually innervate skeletal muscle and spindles. The α-motor neurons are the largest neurons in the spinal cord, with myelinated axons that exit the spinal cord through the ventral roots and travel in peripheral nerves to innervate muscles. Their axons also give off small branches within the spinal cord, called recurrent collaterals, that synapse on inhibitory interneurons, called

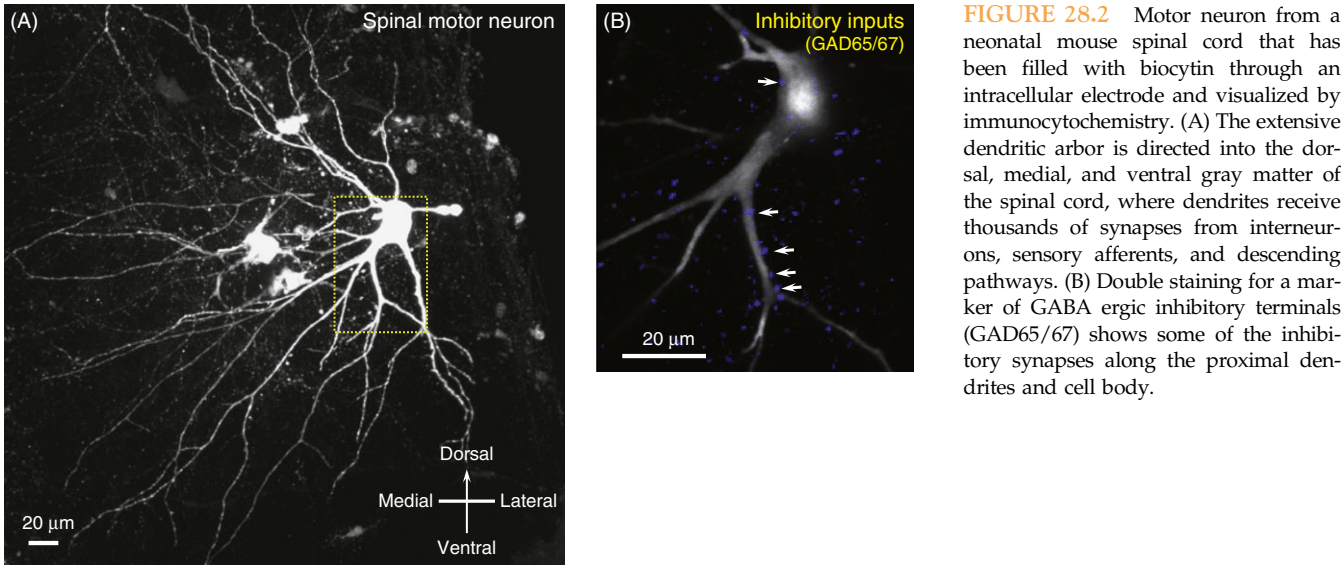


FIGURE 28.2 Motor neuron from a neonatal mouse spinal cord that has been filled with biocytin through an intracellular electrode and visualized by immunocytochemistry. (A) The extensive dendritic arbor is directed into the dorsal, medial, and ventral gray matter of the spinal cord, where dendrites receive thousands of synapses from interneurons, sensory afferents, and descending pathways. (B) Double staining for a marker of GABA ergic inhibitory terminals (GAD65/67) shows some of the inhibitory synapses along the proximal dendrites and cell body.

Renshaw cells (Alvarez & Fyffe, 2007), and other motor neurons. The extensive dendritic tree of motor neurons extends into the dorsal and ventral horns, receiving thousands of synaptic inputs from excitatory and inhibitory neurons (Fig. 28.2). At the neuromuscular junction, the synapse between the motor axon and the muscle fiber, acetylcholine is the transmitter used by mammals, but other transmitters, such as the peptide CGRP or glutamate/aspartate (Mentis et al., 2005; Nishimaru, Restrepo, Ryge, Yanagawa, & Kiehn, 2005), coexist in many motor neurons and may play a particular role during development and at the central synapses formed by recurrent collaterals.

Muscle Contraction

Movement is achieved by the contraction of muscles that are attached to the skeleton or soft tissues, such as muscles producing facial expressions. A skeletal muscle is made up of thousands of individual muscle fibers. Each muscle fiber is a multinucleated cell, a few centimeters long, that contains the contractile proteins, actin and myosin, arranged into thick and thin filaments in repeating units called sarcomeres (Fig. 28.3). The myosin molecule head region contains an ATP ase enzymatic activity that is sensitive to levels of intracellular calcium. Myosin heads are loosely bound to actin at rest, but when intracellular calcium rises, the cross-bridge strengthens, causing myosin to break down ATP. When this occurs, the cross-bridge loosens, and the myosin head swivels along the actin filament to another binding site. The repetition of this swiveling movement in the presence of elevated intracellular calcium and ATP causes thick and thin filaments to slide along each other and shorten the sarcomere length. Shortening of sarcomeres shortens the muscle, or if the muscle is carrying a heavy load, the muscle may only stiffen in an isometric contraction.

The process of muscle contraction begins with an action potential in the motor axon that causes release of acetylcholine at the neuromuscular junction. In mammals, normally every spike in the motor axon produces an action potential in the muscle fiber because an excess amount of acetylcholine is released (Box 28.2). In adult mammals, each skeletal muscle fiber has one neuromuscular junction innervated by one motor neuron. Muscle fiber action potentials lead to the activation of voltage-gated calcium channels in the membrane that interact with calcium release proteins in the sarcoplasmic reticulum, an organelle in the muscle fiber that stores and releases calcium near the sarcomeres. This interaction couples excitation to muscle contraction. Other proteins in the sarcoplasmic reticulum reuptake calcium, leading to relaxation.

Three Basic Muscle Fiber Types

Muscle fibers are classified into types according to their contractile and metabolic properties. Contractile properties refer to the muscle fiber's speed of contraction and relaxation, and the specific force it can produce. Contractile properties are mainly determined by the isoform of myosin that is expressed by the muscle fiber, particularly the myosin heavy chain gene, although the genes for other proteins associated with the sarcomere and the sarcoplasmic reticulum are expressed in fiber-type specific patterns that match the properties of the myosin isoform. The contractile properties of a muscle fiber are classified as slow twitch or fast twitch. Metabolic properties refer to whether aerobic or anaerobic pathways are used to produce energy (ATP) from fuels. Muscle fibers with abundant mitochondria are able to use oxidative (aerobic) metabolism, which

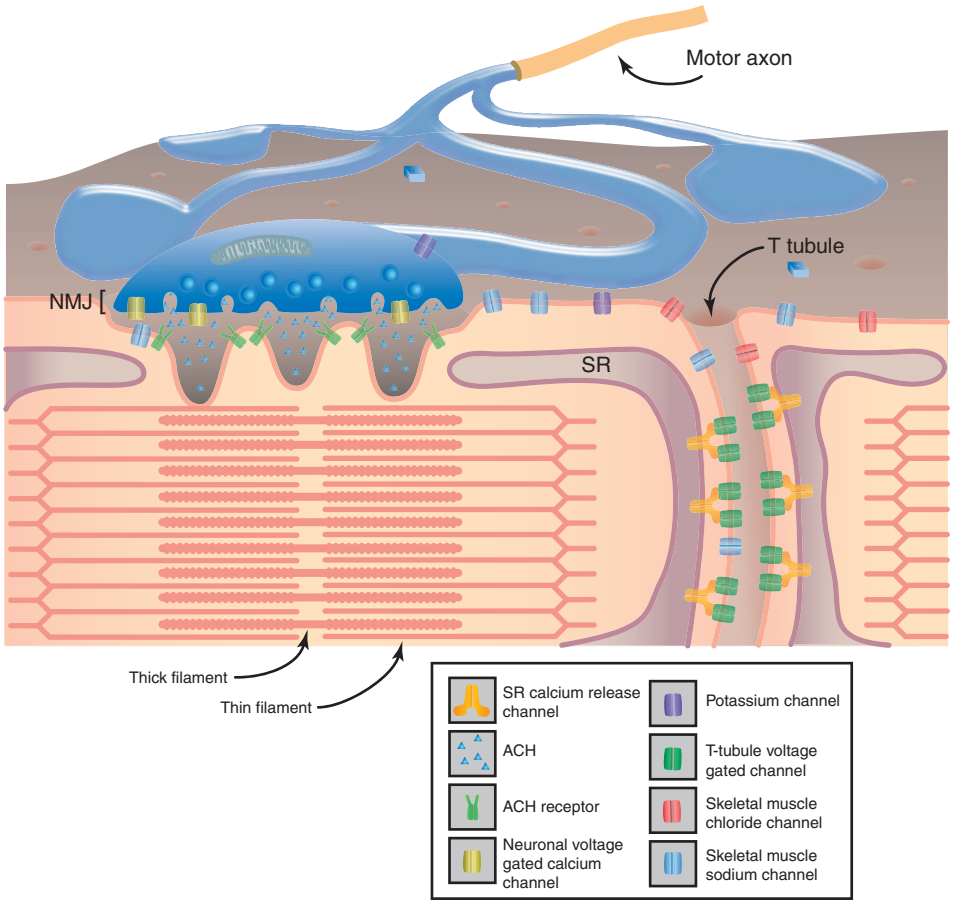


FIGURE 28.3 Diagram of the structure of the muscle showing the relationship between the thick and thin filaments that form the sarcomere, the sarcoplasmic reticulum (SR), and the location of the ion channels of the motor axon, neuromuscular junction, and muscle membrane. *With permission, modified from Cooper and Jan (1999).*

provides relative resistance to fatigue as long as there is a good blood supply. Muscle fibers with few mitochondria rely on stored fuel, such as glycogen, which can more easily be depleted and also produces lactic acid as a byproduct of glycolysis. There are three basic muscle fibers types in adult skeletal muscle. Type I fibers have a slow twitch but have a good capacity for oxidative metabolism. Type IIA fibers have a fast twitch and can use both aerobic and anaerobic metabolism, giving them an intermediate resistance to fatigue. Type IIB fibers produce a fast twitch, but rely mainly on anaerobic glycolysis for their energy. Most muscles contain a mixture of all three fiber types, in proportions that reflect the typical usage of the muscle.

Motor Unit Types

A "motor unit" is defined as one motor neuron and all of the muscle fibers it innervates. In mammals, each muscle fiber is innervated by only one motor neuron, but one motor neuron innervates many muscle fibers as its axon branches in the muscle. However, all muscle fibers of one motor unit are of the same fiber type. Motor units cluster into three basic types based on the three physiological properties of twitch speed, the amount of force produced, and fatigability (Fig. 28.4). Type S motor units are defined by the properties of slow twitch, small amounts of force, and high resistance to fatigue. Type FR motor units are defined by fast twitch, moderately large force, and relative resistance to fatigue. Type FF motor units have fast twitch, produce the highest forces, and fatigue quickly.

The physiological properties of the motor neurons and the muscle fibers they innervate are closely matched. Type S motor neurons innervate Type I fibers. Type S motor neurons have small cell bodies and dendritic arbors, allowing them to be depolarized easily. Afterhyperpolarization following each motor neuron spike promotes a steady rate of firing that allows forces produced by successive twitches of the muscle fibers to summate (Fig. 28.4B). With their steady firing rates and resistance to fatigue, Type S motor units are optimal for the sustained firing needed for postural and tonic movements. Each Type S unit individually contributes only a small amount of force to a muscle's contraction,

BOX 28.2

MYASTHENIA AND MYASTHENIC DISORDERS

Normally every action potential of the motor axon releases more than enough transmitter needed to produce a muscle fiber action potential. This occurs because both the amount of transmitter released from the presynaptic motor axon terminal and the density of AChRs on the postsynaptic muscle fiber exceed that needed to produce an end plate potential (EPP) large enough to initiate a muscle fiber action potential. Genetic or acquired diseases that reduce the safety factor of neuromuscular transmission are called myasthenic disorders (Fig. 28.B1). Myasthenic disorders cause weakness that fluctuates with use of the muscle. The most common is the acquired autoimmune disorder myasthenia gravis, caused by antibodies that bind to AChRs at the neuromuscular junction. The antibodies interfere with ACh binding and lead to internalization of AChRs and, ultimately, the loss of junctional folds. The reduced numbers of functioning AChRs produce a smaller EPP that fails intermittently to trigger a muscle fiber action potential. However, in myasthenia gravis the presynaptic terminal remains normal, such that a short burst of stimulation will produce facilitation of transmitter release, increasing the concentration of ACh in the cleft and transiently improving neurotransmission. Acetylcholinesterase inhibitors provide symptomatic improvement by blocking the breakdown of ACh, allowing a longer effect on residual receptors. Another autoimmune myasthenic disorder, Lambert–Eaton myasthenic syndrome, is caused by antibodies that bind to presynaptic voltage-gated calcium channels. Less ACh is released at the neuromuscular junction, producing small EPPs that fail to trigger action potentials

Congenital myasthenic syndromes (CMS) are extremely rare, but provide important insights into the function of components of the neuromuscular junction. Postsynaptic CMS stem from a deficiency and/or altered kinetic properties of the AChR. More than 40 different mutations in subunits of the AChR have been described. Most mutations decrease subunit expression or prevent subunit assembly or glycosylation, leading to reduced numbers of functioning AChRs and thereby smaller EPPs. In fast channel CMS, mutations produce decreased binding affinity for ACh, leading to a smaller quantal response and miniature EPP amplitude, without reducing the numbers of AChRs. In contrast, the slow channel CMS stem from mutations that prolong the channel opening episodes. Slow channel mutations cause depolarization block due to temporal summation of the prolonged EPPs. In the long term, cationic overloading of the postsynaptic region leads to destruction of the junctional folds and loss of AChRs. Presynaptic CMS are less common. They are caused by mutations that lead to a paucity of synaptic vesicles or defects in ACh resynthesis or vesicular packaging. As a result, fewer vesicles are released or fewer ACh molecules are contained in each vesicle, and EPPs are smaller. Identifying CMS as involving pre- or postsynaptic components and understanding how these alterations affect neuromuscular transmission are necessary for developing rational therapies for these disorders.

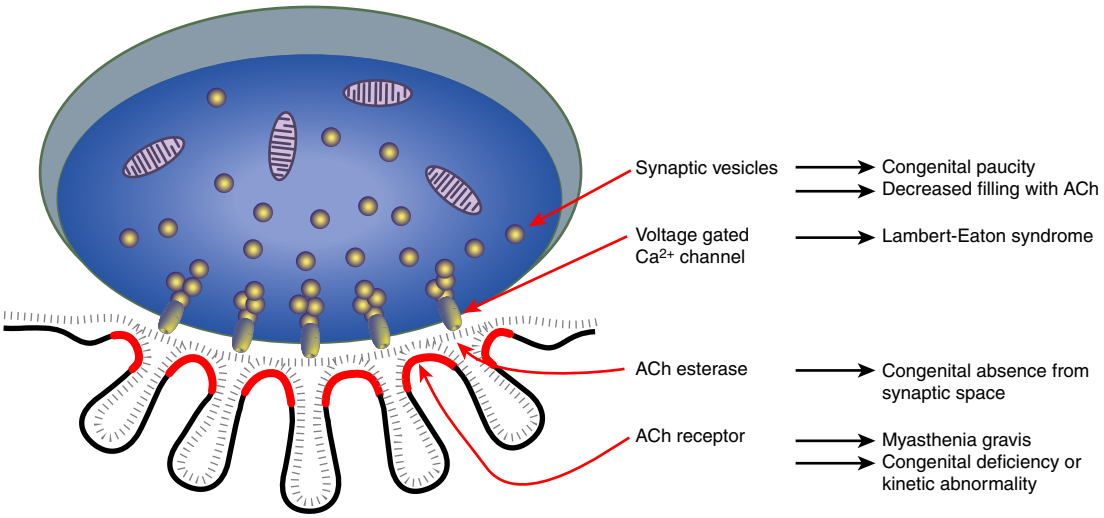


FIGURE 28.B1 Schematic diagram of a neuromuscular junction showing key components subserving neuromuscular transmission and the myasthenic disorders associated with these components.

Mary Kay Floeter

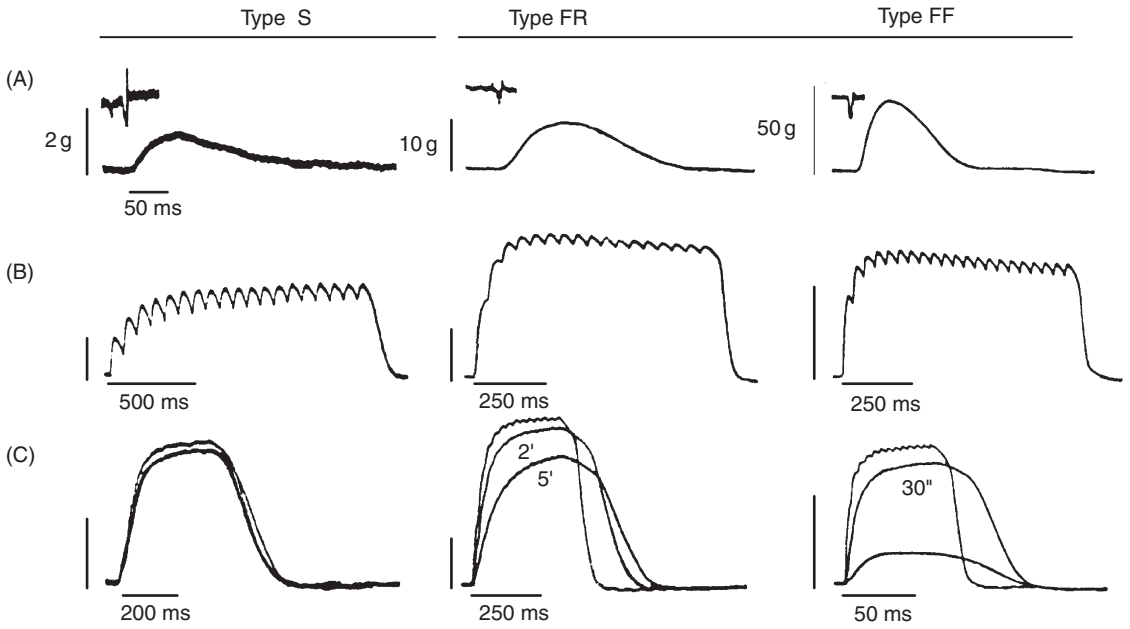


FIGURE 28.4 The three motor unit types (S, FR, and FF) can be defined experimentally by measuring their contractile properties and fatigability. Panels A–C show recordings of muscle force, with insets in A showing recordings of motor neuron action potentials. Note different time and amplitude calibration scales for each of the motor units. (A) Single twitches produced by one action potential of the motor neuron. (B) Maximal force produced by repetitive stimulation of the motor neuron to produce an unfused tetanus. In addition to differences in maximal force, the "sag" property, a dropping off of tension during maintained stimulation, is seen in FR and FF units. (C) Fatigability is demonstrated by a drop in the tension produced by a single twitch after short periods of activation, as noted. Note that S units show little fatigue, whereas FF units fatigue within 30 s. *Reproduced with permission from Burke, Levine, Tsairis, and Zajac (1973).*

but typically Type S units make up a large proportion of the motor units innervating a muscle. In contrast, Type F motor neurons have large cell bodies and dendritic arbors, shorter afterhyperpolarizations and somewhat faster firing rates. Type FR motor neurons innervate Type IIA muscle fibers, and Type FF motor neurons innervate Type IIB muscle fibers. Type FR motor units are optimal for fast, powerful movements. Type FF units are mainly used in movements that require brief bursts of muscle strength. Most muscles contain a mixture of motor unit types, intermingled in a mosaic pattern.

TABLE 28.1 Size-Related Properties of Motor Neurons

Properties that Increase with Size	Properties that Decrease with Size
Diameter of soma and axon	Resistance to fatigue
Conduction velocity	Ia EPSP amplitude
Complexity of axonal collaterals	Input resistance
Membrane area, dendritic extent	Membrane resistance
Rheobase	Time constant
Muscle fiber diameter	Duration of after-hyperpolarization
Maximum force output	Twitch contraction time
	Twitch relaxation time

The Size Principle and Orderly Recruitment

Most movements don't require a muscle to contract at its maximal strength, and only a fraction of the hundred or more motor neurons in the pool that innervates that muscle will need to fire. For most movements, motor units are recruited in an orderly sequence, first described by Henneman (1957), as based on the "size" of the motor neuron. According to the size principle, motor units that produce the smallest amounts of force are the first to begin firing, and motor units that produce larger forces will be progressively recruited as the muscle makes progressively stronger contractions. Many of the anatomical and physiological properties of the motor unit types correlate with measures of size (Table 28.1). These size related properties favor activation of motor unit types in the following order: Type S → Type FR → Type FF. The pool of motor neuron innervating each muscle contains varying proportions of Type S and Type F motor neurons, although generally Type S motor neurons are more abundant.

Many of the properties that produce orderly activation by size are stable and intrinsic to the motor neuron–cell body diameter, for example. Most extrinsic synaptic inputs also reinforce recruitment according to the size principle. This occurs because a synaptic current will produce a larger

postsynaptic potential in a smaller, compact neuron with a higher input resistance. Because most sources of synaptic input are distributed uniformly to a motor pool, EPSP or IPSP amplitudes correlate with the motor neuron size—larger in Type S motor neurons than Type F motor neurons. This means, for example, that stretch of a muscle that activates the primary muscle spindle (Ia) afferent will produce the largest Ia EPSPs in Type S motor neurons, and these will be the first units recruited to fire in the stretch reflex. In the less common situation in which synaptic inputs are not uniformly distributed to every motor neuron in the pool, alternative recruitment sequences are possible.

Stronger muscle contractions are produced not only by recruitment of additional motor units, but also by increasing the firing rates of motor units that are already actively firing. The relative use of recruitment and rate modulation to maintain a desired profile of force varies according to the muscle and the level of force. In some muscles, such as the small hand muscles, several motor units begin firing at low levels of force, and fine gradations of force are controlled by modulating their firing rates. In other muscles, recruitment and rate coding occur together in a more balanced fashion throughout the working force range of the muscle.

Muscle Afferents

Muscles contain two kinds of specialized sensory receptors that provide feedback to the spinal cord and brain. Muscle spindles sense muscle stretch or length, and Golgi tendon organs (GTOs) sense the load upon the muscle (Fig. 28.5). The spindle consists of an encapsulated bundle of modified muscle fibers, called intrafusal fibers, that lie within the muscle in parallel to the skeletal muscle fibers. The spindle receives both sensory and motor innervation. Sensory innervation is supplied by one primary afferent, the Ia fiber, and several secondary, group II afferents. Ia afferents are more sensitive to quick stretches, such as a tendon tap or vibration, and group II afferents maintain a firing rate proportional to the level of sustained stretch. Intrafusal muscle fibers are innervated by γ-motor neurons and β-motor neurons. The motor innervation of intrafusal fibers adjusts the tension of the spindle to compensate for changes in muscle length when a muscle contracts and shortens. When the muscle shortens the spindle can become momentarily slack, causing a pause in spindle afferent firing. However, during most voluntary movements, α and γ-motor neurons are co-activated such that the sensitivity of the spindle is maintained during the muscle shortening. Nevertheless, because γ-motor neurons have a separate innervation of intrafusal muscle fibers, the nervous system can separately adjust the sensitivity of the spindle for different types of movement.

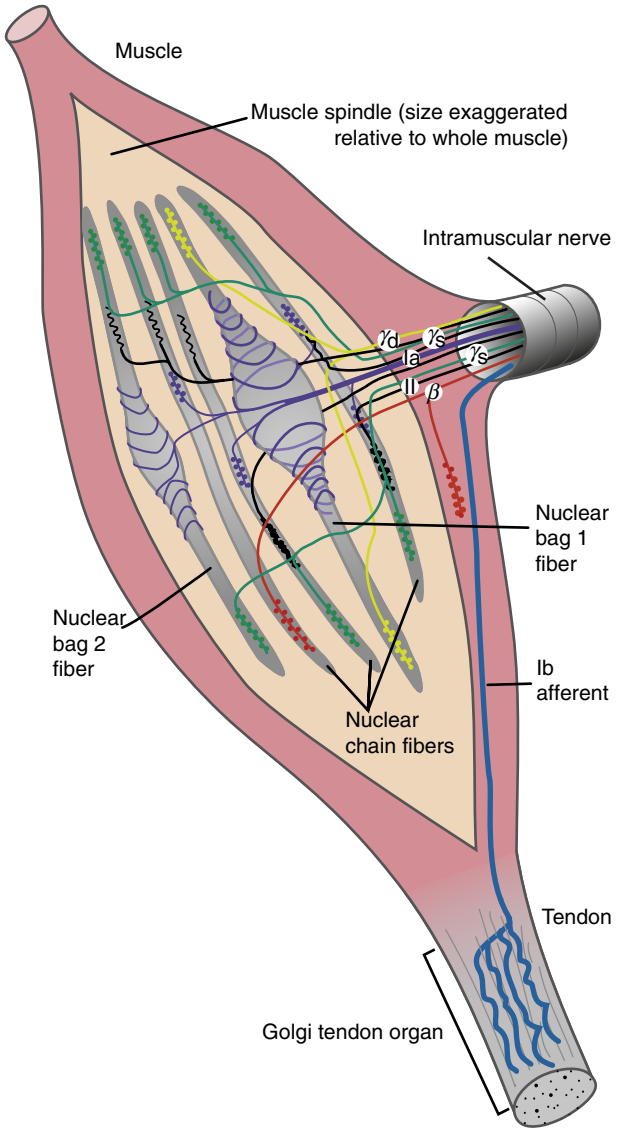


FIGURE 28.5 Structure and innervation of muscle spindles and Golgi tendon organs. The sensory innervation of the spindle is through primary (Ia) spindle afferents that innervate both nuclear bag and nuclear chain intrafusal muscle fibers and secondary (II) afferents that innervate nuclear chain fibers. Motor innervation of the spindle is supplied by static and dynamic γ-motor neurons and by β-motor neurons.

Golgi tendon organs lie at the junction of the muscle and tendon, in series with the muscle. GTOs are sensitive to the forces generated by muscle contraction against a load. Each GTO senses the force generated by a small number of motor units that insert onto a local region of the tendon. The axons of the GTOs are called Ib afferents. Like Ia afferents, they are large myelinated axons with fast conduction velocities. Unlike spindles, GTOs do not have motor innervation.

Muscles are innervated by smaller diameter and unmyelinated afferents, called group III and IV afferents, that respond to extracellular substances, such as the acidic substances released during fatiguing exercise. These slowly conducting afferents relay the sense of muscle pain and soreness and produce reflex actions with a relatively slow time course.

Stretch Reflexes

The muscle stretch reflex plays an important role in providing feedback during movements. The stretch reflex consists of a monosynaptic response from the direct connection between Ia afferents and motor neurons, that may be followed by polysynaptic reflex activity. At rest, the stretch reflex can be obtained by a quick stretch, such as by tapping on a muscle tendon, which produces a burst of action potentials in the Ia spindle afferents (Fig. 28.6). In the spinal cord, the Ia afferents synapse on the motor neurons of the muscle containing the spindle (the homonymous motor neurons) and, to a lesser extent, on motor neurons that innervate other muscles (heteronymous motor neurons). The rapid volley of glutamatergic Ia EPSPs caused by a quick stretch can summate in the motor neuron, and if sufficient depolarization occurs, an action potential occurs that produces contraction of the muscle, with a jerk of the limb.

If the stretch reflex is elicited during movement, the Ia EPSPs will summate with inputs from other sources such as CPGs and descending inputs. Stretch during the phase of movement when the muscle is active will add to other sources of motor neuron depolarization, and the reflex motor response will enhance the ongoing contraction. Stretch of the muscle when it is out of its active phase will produce a smaller motor response, if any. In this way, the spatial summation of postsynaptic potentials contributes to phase modulation of the stretch reflex. Additionally, there is modulation of the stretch reflex through presynaptic inhibition of the Ia afferents during movement (Rudomin, 2009). This presynaptic inhibition is produced by GABAergic spinal interneurons that synapse on the Ia afferent terminals, forming axo-axonic synapses (Betley et al., 2009). Presynaptic inhibition reduces the release of glutamate from Ia afferents, leading to smaller Ia EPSPs in the motor neuron, but it does not affect the size of other inputs to the motor neuron. The GABAergic interneurons mediating presynaptic inhibition are strongly activated by antagonist Ia afferents, that is, from spindles of muscles that produce the opposite movement, as well as by CPGs and descending inputs. Thus, presynaptic inhibition of a muscle's stretch reflex is strongest when the muscle contraction is out of phase with the movement cycle, when its antagonist muscle is contracting.

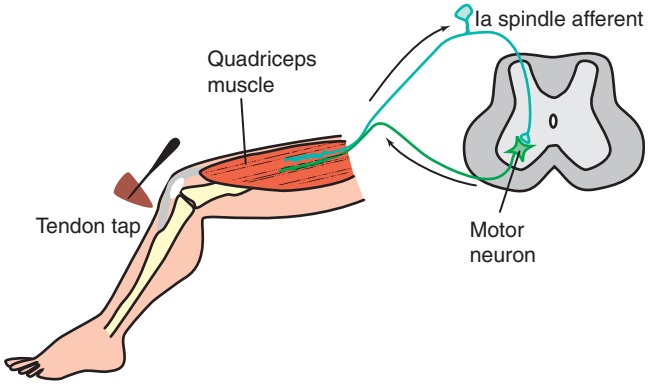


FIGURE 28.6 Basic circuitry underlying the knee-jerk stretch reflex. Ia spindle afferents from the quadriceps muscle make monosynaptic, excitatory connections on α-motor neurons that innervate the quadriceps.

Summary

Motor neurons are highly diverse in terms of their morphology, connectivity, and functional properties and differ significantly in their response to disease. There are two main types of motor neuron disease, amyotrophic lateral sclerosis (ALS) and spinal muscular atrophy (Box 28.3). Motor neurons innervating different muscles have recently been shown to degenerate in a preferential manner.

Muscles are innervated by motor neurons in the ventral horn of the spinal cord. The motor unit consists of one motor neuron and all the muscle fibers it innervates. Physiological properties of each motor neuron match the contractile and metabolic properties of its muscle fibers. Type S motor units have a slow twitch, high aerobic capabilities, and fatigue resistance. These smaller motor units are the first recruited for contraction and are heavily used for slow, sustained postural movements. Type F motor units innervate muscle fibers with a fast twitch and are recruited for more forceful or ballistic movements. Type FR motor units use aerobic and anaerobic metabolism and are relatively resistant to fatigue. Type FF motor units use anaerobic metabolism and fatigue quickly. In the pool of motor neurons innervating a muscle, motor neurons are recruited in an orderly fashion, beginning with Type S units that generate small amounts of force, followed by units that generate greater forces. As the muscle contracts, changes in its length are detected by spindles, specialized receptors that lie within the muscle. The load on the muscle is detected by a second specialized receptor, the Golgi tendon organ. These sensory receptors of the muscle provide feedback to the spinal cord allowing reflex changes in drive to the muscles.

BOX 28.3

MOTOR NEURON DISEASES

There are several neurological disorders that selectively affect motor neurons. Spinal muscular atrophy (SMA) is a hereditary disorder in which motor neurons develop normally but begin to degenerate not long after birth. Autosomal recessive SMA was described by neurologists long before its cause was known, and several different clinical phenotypes were described. SMA causing weakness in infancy was called Werdnig-Hoffman disease; SMA causing weakness later in childhood was called Kugelberg-Welander disease. Both forms of SMA are caused by a mutation in a gene called "survival of motor neuron" or SMN gene. The SMN protein is widely expressed in neurons and has a role in processing mRNAs. It is not known why motor neurons are more vulnerable to mutations of SMN than other cells. In an X-linked SMA, called Kennedy's disease or spinal and bulbar muscular atrophy (SBMA), weakness begins in adult life in affected males. Motor neurons innervating facial and jaw muscles are often affected early on. The mutation in SBMA is a triplet repeat expansion in the androgen receptor gene that leads to expression of an abnormal androgen receptor. The androgen receptor is expressed in motor neurons, and over time accumulation of the abnormal receptor leads to degeneration of the motor neuron.

Amyotrophic lateral sclerosis (ALS), also known as Lou Gehrig's disease, is a disease with degeneration not only of the motor neurons, but also of corticospinal neurons of the motor cortex. In most patients with ALS, the disease is not hereditary and its cause is unknown. There are a few hereditary forms of ALS that make up a small percentage of all patients with ALS. Nevertheless, the identification of familial ALS (FALS) genes has provided ideas about pathways that may be responsible for causing sporadic ALS. The first gene mutation identified in FALS encoded the enzyme superoxide dismutase, SOD1. Studies using transgenic mice expressing the mutant SOD1 gene showed that FALS was not caused by loss of enzyme activity, but by abnormal properties of the mutant SOD1 protein. Studies expressing mutant SOD1 in different cell types are revealing that glial and inflammatory cells may also contribute to motor neuron degeneration in FALS (Boillee, Vande Velde, & Cleveland, 2006). The ability to manipulate the expression of mutant genes in animal models promises to provide an invaluable contribution to finding causes and treatment of ALS.

Mary Kay Floeter

SPINAL INTERNEURON NETWORKS

So far, this chapter has discussed how the output of the spinal cord produces muscle actions through the motor neurons. This section will focus on spinal interneurons—neurons with a cell body in the spinal cord and an axon that synapses on another spinal neuron. Spinal interneurons have three important functions for motor control: they relay sensory inputs from the periphery that may modulate the motor output, they relay as well as modulate signals from the brain via descending pathways, and they form networks that produce patterned rhythmic activity. It is important to note that some interneurons may participate in all three functions, switching from one role to another during specific movements or carrying on two roles simultaneously.

Historically, interneurons were identified and named by their axonal projections, the location of their soma, their neurotransmitter, or by a physiological effect or reflex that they mediated. Later, populations of neurons were identified by their pattern of activity during movements, visualized using activity-dependent dyes. More recently, molecular biology tools have enabled scientists to identify interneurons based on their gene expression pattern and cell lineage. Molecular markers have been used to identify several classes of interneurons and motor neurons. Four distinct populations of interneurons, designated as V0, V1, V2, and V3, have been identified in the ventral spinal cord by their developmental pattern of expression of transcription factors, proteins that regulate gene expression (Jessell, 2000; Goulding & Pfaff, 2005; Grillner & Jessell, 2009). The "V" denotes the *ventral* location, and each subclass appears at a distinct position along the dorso-ventral axis (0 most dorsal, 3 most ventral) of the ventral spinal cord (Fig. 28.7). The cellular identities of spinal interneurons are further refined by the emergence of combinations of transcription factors that are expressed in these neurons (Goulding 2009; Alaynick, Jessell, & Pfaff, 2011).

Models for Studying the Locomotor CPG Network

At the beginning of the chapter, we noted that the CPG network needed to produce three fundamental features of locomotion: generation of a cyclical rhythm, coordinated alternation between antagonist flexor and extensor muscles, and coordination between limbs. Figuring out what makes up a CPG—the neurons involved, the connections between

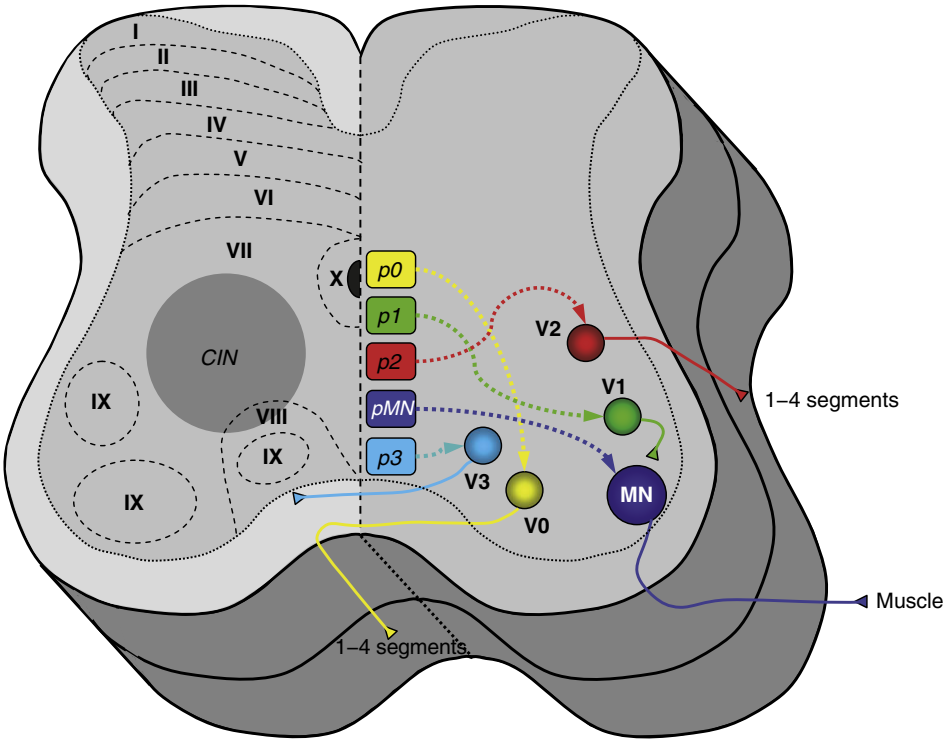


FIGURE 28.7 Cross-sectional schematic of the spinal cord depicting on the right side the origin of interneuron and motor neuron populations defined in developmental studies using molecular markers for cell identification. The ventral interneurons (V), derived from progenitors (p), are marked as V0–V3 based on the distinct position along the dorso-ventral axis (0 most dorsal, 3 most ventral) of the ventral horn. The left side of the spinal cord depicts the general location of commissural interneurons (CINs) and the Rexed's laminae. CINs are characterized based on the projection of their axons, as ascending, descending, or both.

these neurons, the synaptic interactions, and the contributions from cellular ion channels—is challenging. To tackle this problem, researchers construct models that can be tested and refined by experimental observations. As noted earlier, Brown (1911) first proposed a model with mutually inhibitory half-centers driving flexors and extensors. Grillner (1985) proposed three important modifications: (1) that at each joint, movement in one direction was produced by a "burst generator" unit, a half-center capable of independent rhythmic activity; (2) as in Brown's model, that flexion and extension at each joint were controlled by reciprocally connected burst generators to form a unit CPG; and (3) that multijoint coordination was achieved by loose or adjustable coupling between the mosaic of unit CPGs controlling each of the joints (Fig. 28.8).

Experimental findings have led to further refinements of the proposed organization of the CPG network. For example, the rhythm and timing of the network may be controlled separately from the patterning of the network output. In studies of locomotor rhythms in the neonatal rodent spinal cord, the CPG appears to be distributed among several spinal segments. Studies in the isolated neonatal rodent spinal cord preparation maintained *in vitro* (Fig. 28.9) show a difference in the rhythmogenic capacity between the rostral (upper) and the caudal (lower) segments of the lumbar spinal cord. The upper lumbar (L1–L3 in rodents) segments seem to be most important for rhythm generation compared to caudal lumbar segments (Cowley & Schmidt, 1997; Tresch & Kiehn, 1999). In experiments using calcium dyes to

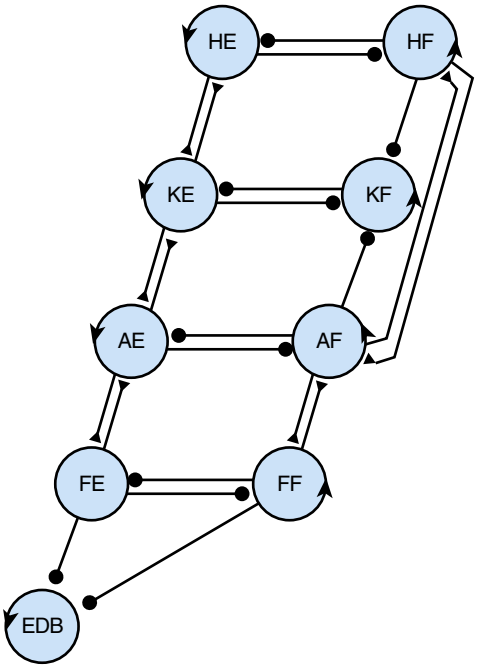


FIGURE 28.8 Scheme for multijoint coordination within the leg using a mosaic of "unit CPGs." Burst generators produce movement in one direction, extension or flexion, at the hip (HE, HF), knee (KE, KF), ankle (AE, AF), or foot (FE, FF), with a separate generator for the muscle that lifts the toes (EDB). Burst generators for flexion and extension at each joint are mutually inhibitory, forming a unit CPG that produces alternation. The strength of excitatory (triangles) and inhibitory (circles) connections between unit CPGs can vary to create a variety of different gaits. *Reproduced with permission from Grillner (1985).*

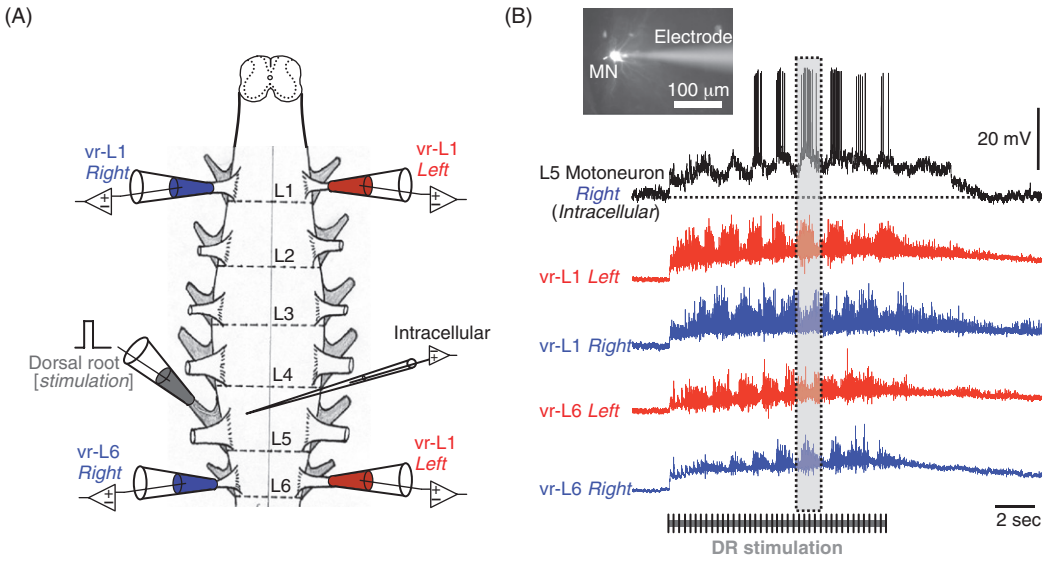


FIGURE 28.9 (A) Drawing of an isolated neonatal mouse spinal cord used in *in vitro* experiments. Suction electrodes are placed in the ventral lumbar (L) roots to record motor neuron activity (blue and red) or in the dorsal root (gray) to stimulate sensory afferents. Glass electrodes are used to record intracellularly from individual neurons in the spinal cord. (B) The image shows a motor neuron (MN) filled with a fluorescent dye and visualized with a microscope during intracellular recording. Sustained electrical stimulation of sensory afferents can evoke rhythmic activity in motor neurons (intracellular record) characterized by rhythmic membrane depolarization which reaches the threshold for action potentials. This stimulation also results in alternating rhythmic activity between the left and right side of the spinal cord [compare the activity in L1-left (red) and L1-right (blue)]. It also produces alternating activity between the upper (L1) and lower (L6) segments of the spinal cord [compare the 2 red traces or the 2 blue traces], as highlighted in the dotted box. The duration of afferent stimulation is indicated by the graded bar.

visualize motor neuron activity, rhythmic signals from different parts of the lumbar (L1, L2) and sacral (S1–S3) segments rose, peaked, and decayed in a rostrocaudal sequence, giving rise to a rostrocaudal "wave" of motor neuron activation during each locomotor cycle (Bonnot, Whelan, Mentis, & O'Donovan, 2002). The importance of the upper and mid-lumbar cord for rhythm generation has also been observed in the spinalized cat and is consistent with case reports in humans noting that irritative lesions of the upper lumbar cord produce stepping movements.

Excitatory Interneurons Responsible for the Rhythm

Maintaining rhythmic bursting in an interneuronal network requires a source of excitatory drive. Although this drive could be provided by tonic excitation, such as from descending pathways, a more robust rhythm is possible if there are excitatory cells within the network that have intrinsic properties that allow rhythmic and repetitive firing (Box 28.4). A few classes of rhythmically active excitatory interneurons have to date been identified in the mouse spinal cord and implicated in central pattern generation. The first type of interneurons is a population of V3-derived interneurons expressing the transcription factor Sim1 (Zhang et al., 2008). These interneurons release glutamate and target motor neurons, as well as V1 and V2 interneurons. The second type of interneurons is a subpopulation of V2 interneurons, termed V2a, which express the transcription factor *Chx10* (Crone et al., 2008; Zhong et al., 2010). The *Chx10*+ interneurons release glutamate and also contact motor neurons directly. Another candidate population of excitatory interneurons was first identified in mice with deletions in the genes for two axon guidance molecules, *EphA4* and *ephrinB3* (Kullander et al., 2003). These mice had an abnormal hopping pattern of gait. Subsequent study of interneurons expressing *EphA4* receptors showed that they were glutamatergic and rhythmically active during locomotor behavior (Butt, Lundfald, & Kiehn, 2005). Most of the interneurons were "last order" interneurons—that is, interneurons that synapse directly on motor neurons. Finally, another interneuron that is rhythmically active during fictive locomotion using isolated spinal cords of mice is the interneuron that expresses the transcription factor Hb9, which is also expressed in the motor neuron lineage during development (Hinckley, Hartley, Wu, Todd, & Ziskind-Conhaim, 2005; Wilson et al., 2005).

Although these interneurons have been proposed as candidates forming the rhythm generating circuitry of the CPG, it is important to note that more work needs to be done to prove their essential role in rhythmogenesis. Not all rhythmically active neurons are components of the CPG network. Some may receive rhythmic output from CPG neurons. Causality needs to be established

BOX 28.4

NONLINEAR PROPERTIES OF MOTOR NEURONS AND CPG INTERNEURONS

Experimental evidence from animal work indicates that motor neurons can participate actively in generating the rhythmic pattern observed during locomotion. Some motor neuron properties can be turned *ON* during locomotion due to synaptic inputs which in turn, produce nonlinear input–output relationships.

For example, there is a locomotor-dependent reduction in the threshold for spike generation, resulting in higher motor neuronal excitability. Two other motor neuronal properties susceptible to interact significantly with afferent inputs are the membrane potential oscillations known as *locomotor-drive-potentials* (LDP) and *plateau potentials*. LDPs are membrane potentials in motor neurons, in which a wave of depolarization during the active phase alternates with a wave of hyperpolarization during the antagonist phase of locomotion. Experiments in the cat have shown that the depolarizing phase is due to an excitatory synaptic drive, while the hyperpolarizing phase is due to an inhibitory drive.

Plateau potentials are a behavior of the membrane potential in which the potential jumps between two states (bistable behavior). This motor neuron behavior has been observed in animals such as the turtle, mouse, and cat. This behavior is mediated mostly by a persistent inward current that at least in part is mediated by calcium channels. Ionic currents responsible for the generation of plateau potentials in motor neurons are found predominantly in dendrites because they possess voltage-sensitive L-type calcium conductances. These currents are turned *ON* by some neuromodulators such as serotonin or noradrenalin. Data from animal experiments have shown that synaptic inputs on dendrites can recruit plateau potentials and evoke sustained discharge. It is possible therefore that dendritic activation can amplify the effect of synaptic inputs from either sensory or descending afferents. Although the role of plateau potentials is still unclear during normal development, one hypothesis is that motor tasks like posture or stepping recruit fatigue-resistant motor units readily with sustained discharge produced by persistent inward currents without continuous network inputs.

From animal experiments there are some preliminary data that indicate that plateau potentials and sustained discharge are also found in some spinal interneurons. As in motor neurons, these plateaus are facilitated by serotonin and noradrenalin. In a similar fashion as in motor neurons, plateau potentials activated in interneurons by neuromodulators released during locomotion may act as an amplification step of synaptic inputs in particular dendrites. Because interneurons integrate sensory and supraspinal information, this mechanism may be very effective in the transmission of inputs relevant for the generation of motor neuronal activities.

George Mentis

experimentally by temporarily and reversibly removing the participation of a population of interneurons during fictive locomotion. Strategies using genetic markers that identify specific populations of interneurons to drive selective expression of reversible inhibitors of activity provide a potential tool to test the contribution of distinct interneuron populations to CPG networks (Lechner, Lein, & Callaway, 2002).

Reciprocal Inhibition between Flexor and Extensor Muscles

Alternation between flexor and extensor muscles that act across the same joint is another fundamental feature of locomotion, as well as of many other limb movements. There are several spinal interneuron circuits that act in different ways to produce reciprocal inhibition between antagonist muscles. Presynaptic inhibition of antagonist Ia afferents by GABAergic interneurons, described earlier, prevents spindle afferents from activating a muscle when it is stretched by the contraction of its antagonist. Another interneuron, the glycinergic Ia inhibitory interneuron (IaIN), produces postsynaptic reciprocal inhibition. The IaIN was so named because it is activated by the muscle's Ia afferents and inhibits motor neurons innervating the antagonist muscle (Wang, Li, Goulding, & Frank, 2008). The IaIN synapses on the soma and main dendrites of motor neurons, producing a fast, potent IPSP that lasts only for several milliseconds. The IaIN develops from the V1 cell lineage, and can be recognized by its expression of the transcription factor *En1* (Siembab et al., 2010). Almost every input that excites a motor neuron will also excite the IaINs that project to its antagonist's motor neurons. For this reason, voluntary contraction of a muscle will produce inhibition of its antagonist muscle, and, likewise, the stretch reflex of a muscle will produce inhibition of its antagonist. The excitability of IaINs is determined by their many inputs, and the strength of reciprocal inhibition

can be varied to allow movements that use co-contraction. IaINs are rhythmically active during locomotion, even without feedback from muscle stretch (Pratt & Jordan, 1987). This observation led to the proposal that IaINs were involved in producing the flexor-extensor alternation of the locomotor CPG. However, this proposal is not supported by more recent studies using neonatal mice in which the V1 lineage neurons were reversibly inactivated. In this study, the locomotor rhythms became slower, but still exhibited alternation of flexors and extensors (Gosgnach et al., 2006).

Interneurons Involved in Left–Right Coordination

Interlimb coordination is the third fundamental feature of locomotion. To date, the leading candidate interneurons for coordinating activity between the limbs during locomotion are several populations of commissural interneurons (Fig. 28.10). Commissural interneurons have axons that cross the midline of the spinal cord at the level of the ventral commissure. They can be subdivided into three classes based on the direction of their axon: descending, ascending, or both (bifurcated axon). Descending commissural interneurons have been extensively studied in the rodent spinal cord during early development. Their axons project at least two spinal segments away from their soma to make contacts with other ventral horn interneurons and motor neurons. Some ascending commissural interneurons are cholinergic and are thought to transmit information to the upper parts of the spinal cord. Many of the commissural interneurons arise from the V0 and V3 cell lineage. The V0 lineage gives rise to ascending commissural interneurons, whereas the V3 lineage comprises a mixed population of commissural as well as ipsilaterally projecting neurons. In transgenic mice lacking the V0 commissural interneurons, the coordination between left and right legs was disrupted during locomotor rhythms, with episodic synchronous activation between sides (Lanuza, Gosgnach, Pierani, Jessell, & Goulding, 2004). These results suggest that this class of ascending commissural interneurons contributes, in part, to side-to-side interlimb coordination.

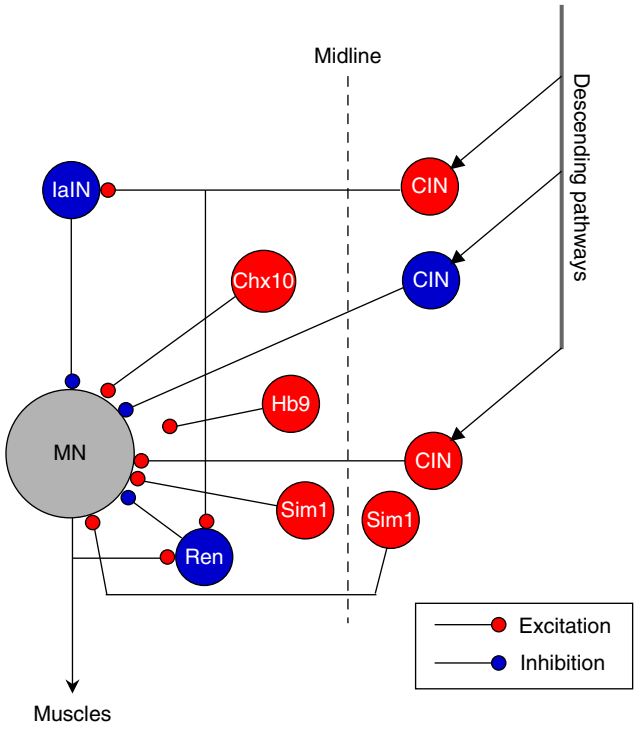


FIGURE 28.10 Schematic model with some of the known candidate interneurons that are involved in the modulation of the mammalian locomotor CPG. Excitatory interneurons (red) with rhythmic activity include classes of commissural interneurons (CIN) and the interneurons identified by expression of Sim1, Chx10, and Hb9 transcription factors. Inhibitory interneurons (blue) that participate in termination of excitatory bursts include the Renshaw (Ren), the Ia interneuron (IaIN), and some commissural interneurons (CIN). *Modified from Grillner (2006).*

Recurrent (Renshaw) Inhibition

The recurrent collaterals of motor neurons synapse on a population of inhibitory interneurons in the ventral horn, called Renshaw cells (Renshaw, 1946). Renshaw cells use glycine or GABA as their transmitter and during development may contain both transmitters. Renshaw cells project back to homonymous motor neurons and synergist motor neurons in adjacent segments. They provide a negative feedback reflex called recurrent inhibition that limits the firing of motor neurons (Eccles, Fatt, & Koketsu, 1954). During locomotor rhythms, Renshaw cells are rhythmically active, as expected, since they receive rhythmic synaptic inputs from motor neurons. Their role in voluntary movements is uncertain, but one possibility is that they enhance the contrast between the active motor neurons and quiescent synergist motor neurons in adjacent segments. Renshaw cells arise from the V1 cell lineage, as do IaINs. During development, Renshaw cells, like IaINs, receive input from Ia afferents (Mentis, Siembab, Zerda, O'Donova, & Alvarez, 2006). In the studies in neonatal mice, mentioned above, reversible inactivation of the V1 lineage neurons that produced slowing of locomotor rhythms included Renshaw cells and IaINs (Gosgnach et al., 2006).

Summary

In addition to serving as relays for sensory and descending pathways, interneurons of the spinal cord interconnect to form CPG networks that can produce rhythmic firing. Several models of the CPG network have been proposed to guide experimenters. There has been significant progress in unraveling the

interneuronal network that forms the spinal central pattern generator for locomotion. New experimental techniques have allowed the identification in neonatal rodents of interneurons that are candidates for mediating some of the fundamental properties of locomotion: rhythmicity, alternation of antagonist muscles, and interlimb coordination.

DESCENDING CONTROL OF SPINAL CIRCUITS

Neural circuits in the brainstem, cerebellum, and cerebral cortex are activated with locomotion and other movements. For locomotion, the cerebral cortex not only formulates conscious decisions about initiating and maintaining walking, but plays a role in adaptations of gait, particularly visually guided walking. Recordings from corticospinal neurons in cats as they stepped over obstacles placed in their field of vision showed that many corticospinal neurons changed their firing rates, often with a timing appropriate for driving the altered contraction patterns of specific muscles (Drew, Prentice, & Schepens, 2004). Corticospinal input, affecting motor neurons directly or through spinal CPG interneurons, can alter the limb's trajectory with a high degree of precision, in a similar manner to a goal-directed reaching movement. In people, clinical observations and experimental studies suggest that supraspinal regions play a more active role in locomotor control than in lower mammals (Box 28.5).

BOX 28.5

NEURAL CONTROL OF HUMAN WALKING

Human walking is a complex behavior that is based on integration of activity from descending supraspinal motor commands, spinal neuronal circuitries, and sensory feedback. Accumulating evidence suggests that humans, as well as other species, have a network in the spinal cord that is capable of generating basic rhythmic walking activity. Rhythmic leg movements can be induced by epidural electrical stimulation after a clinically complete spinal cord injury (SCI). Spontaneous involuntary rhythmic leg movements have also been reported after a clinically complete and incomplete SCI. However, even though this evidence suggests the existence of a CPG network in humans, in all these cases sensory input likely contributed to rhythm generation. Furthermore, clinically complete injuries are not always anatomically complete. As a result, the evidence for existence of a CPG in humans remains indirect.

Noninvasive electrophysiological techniques have provided information about how certain brain structures contribute to human walking. For example, the involvement of the primary motor cortex, where the corticospinal tract originates, has been demonstrated, in part, with transcranial magnetic stimulation (TMS). Several groups have found changes in the size of motor evoked potentials (MEP) in lower limb muscles elicited by TMS during walking, showing that transmission in the corticospinal tract is modulated during the gait cycle. However, TMS of the motor cortex activates cells with monosynaptic or polysynaptic connections to the spinal motor neurons. The size of the MEPs elicited by TMS reflects not only cortical excitability changes, but also changes at a sub-cortical level. To more precisely identify the various sites activated by the magnetic stimulus, Petersen, Christensen and Nielson (1998) tested the effect of TMS on the H-reflex, an electrical homologue of the monosynaptic stretch reflex, during walking. They found that at least part of the MEP modulation was caused by changes in the excitability of corticospinal cells with direct monosynaptic projections to the spinal motor neurons. They also found that TMS over the motor cortex at intensities below the threshold for activating spinal motor neurons depressed ongoing muscle activity in the tibialis anterior muscle of the leg during walking. This effect was thought to be caused by activation of cortical inhibitory interneurons targeting cortical pyramidal tract neurons (Petersen et al., 2001).

During human walking, sensory feedback plays an important role in driving the active motor neurons and providing error signals to the brain that can be used to adapt and update the gait pattern. In healthy humans, a study in which a plantarflexion perturbation was induced during the stance phase of the gait cycle provides clear evidence for this. This experiment showed a drop in electromyographic (EMG) activity in the plantarflexor muscles, even when the common peroneal nerve that innervates the ankle dorsiflexor muscles was blocked by local anesthesia (Sinkjaer, Anderson, Ladouceur, Christenson, & Nielson, 2000). These results demonstrated that sensory feedback from plantarflexor muscle contributed to the active drive of motor neurons during walking. Additional information about the contribution of sensory feedback to the control of locomotion has been gathered from studies that investigate changes in EMG responses in lower limb muscles after stimulation of cutaneous and proprioceptive afferents during walking. The amplitudes of the EMG responses to electrical stimulation are dependent on the phase (or time) of stimulation during

BOX 28.5 *(cont'd)*

the gait cycle. For example, electrical stimulation of the human sural nerve results in facilitation of the ankle flexor muscle tibialis anterior during early swing, but leads to suppression when delivered during late swing (reflex reversal). It has been proposed that the phase-dependent modulation of EMG responses may have functional relevance at various times in the step cycle. For example, a flexor reflex is appropriate at the onset of the swing phase when the leg is flexed, but the same reflex is not convenient at the end of the swing phase when the foot is ready to take up body weight. To some extent this modulation is mediated by supraspinal centers. To examine this possibility, a combination of cutaneous input and activation of supraspinal pathways by TMS has been used during the gait cycle. Available data suggest that long-latency reflexes elicited by cutaneous electrical stimulation are in part transcortically meditated.

In summary, in recent years a combination of neurophysiological techniques has provided important insights that have improved our understanding about the neural control of human walking. This behavior is based on the integration of activities from spinal circuitry, sensory feedback, and descending motor commands.

Monica Perez

Brainstem Control of Spinal Interneuron Networks

The brain transmits signals for movement to the spinal cord through several descending fiber tracts: corticospinal, vestibulospinal, and reticulospinal, as described in the following chapter. For locomotion, supraspinal regions are thought to activate spinal cord CPGs through brainstem relays. In decerebrate cats, electrical stimulation of several areas of the brain can elicit stepping. These include the diencephalon and cerebellum, but the most critical region is an area in the brainstem near the pedunculopontine nucleus known as the mesencephalic locomotor region (MLR). The MLR receives inputs from several forebrain regions, including the basal ganglia. The MLR does not connect directly to spinal CPG neurons, but synapses on reticulospinal neurons in the pons and medulla that receive a wide variety of sensory and descending inputs. Some of the reticulospinal axons in the pontomedullary formation travel in the ventral and ventrolateral tracts (funiculi) of the spinal cord. Lesions of these funiculi disrupt the ability to initiate and maintain locomotion. Sparing of at least some fibers in the ventrolateral funiculus markedly improves locomotor recovery after spinal lesions in cats (Rossignol, Drew, Brustein, & Jaing, 1999) and could be critical for the recovery of locomotor movements after spinal injury in humans.

In addition to their role in triggering rapid changes in locomotion and regulating postural balance, descending fiber systems also modulate physiological properties of spinal neurons. For instance, the plateau potentials and bistability in motor neurons (Box 28.4) depend on tonic activity in descending serotonergic pathways. Stimulation of the dorsolateral funiculus of the spinal cord in turtles, an area which is thought to contain descending pathways from the raphe nucleus, increased the excitability of motor neurons and promoted production of plateau potentials.

SENSORY MODULATION

Although spinal CPGs can generate the basic pattern of neural activity that organizes muscle contractions into a recognizable movement such as locomotion, the final pattern is fine-tuned by sensory inputs as well as descending systems. As noted earlier, most sensory inputs affect motor neurons through polysynaptic, interneuronal circuits. The interneurons that process sensory information typically integrate sensory signals with inputs from a variety of sources, including CPG networks and descending pathways. As a result the reflex motor response can vary depending on the phase of the movement cycle, or it may only occur when the animal is in the "state" of moving, or it may even be reversed between resting and active states. Earlier we discussed the muscle sensory receptors and some of the reflex changes they mediate. This last section will present a sampling of other sensory reflex pathways.

Flexor Reflexes

A wide variety of sensory stimuli can evoke reflex flexion of the limb. Flexor reflexes may have several roles: they serve a protective function, reflexively withdrawing the limb from potentially harmful stimuli, and they may be used in movements that require coordinated multijoint flexor movements. Usually the reflex consists of contraction of muscles that flex several joints coupled with relaxation of muscles that extend those joints. Extension of the contralateral limb may also occur. Flexor reflexes can be elicited by group II and III muscle afferents, joint afferents, and high- and low-threshold cutaneous afferents. Because these afferents elicit a common action, they are sometimes called "flexor reflex afferents" (FRAs). The muscles activated by the sensory stimulus respond with short and long latency bursts of contraction. These contraction components are thought to be generated by distinct

interneuron circuits. The late components can be differentially enhanced by treatment with L-dopa in spinalized animals. It has been suggested that interneurons mediating the late flexor reflex could function as a half-center for flexion, as proposed in models of the locomotor CPG.

Cutaneous Reflexes

In contrast to the widely distributed flexor reflex, there are a number of reflexes in which stimulation of small areas of skin evokes a reflex limited to a few muscles. One example of a local cutaneous reflex is the stumbling corrective reaction. When the skin on the top of the foot is stimulated during walking, the foot is lifted higher, as if to step over an obstacle. The stumbling corrective reaction is phase dependent, occurring only when the limb is in the swing phase of the step cycle. This and several other local cutaneous reflexes are oligosynaptic, involving only a few interneurons. The last-order interneurons in these cutaneous reflexes connect with relatively limited sets of motor neurons. It has been proposed that such interneurons offer a means to control or alter activity of specific muscles during movement without involving the interneurons of pattern generating circuitry.

Summary

Several areas of the brain give rise to descending pathways that connect to spinal interneurons and, to varying extents in different species, to motor neurons. Areas of the brainstem are critical for relaying descending commands for locomotion. Volitional movements also seem to employ spinal interneuron circuits for organizing routine patterns of muscle contraction, within the context of a planned goal. Sensory feedback from skin, joints, and most muscle afferents provide fine-tuning of basic movement patterns. The reflex motor responses evoked by most sensory afferents are relayed through interneurons.

References

Alaynick, W. A., Jessell, T. M., & Pfaff, S. L. (2011). SnapShot: Spinal cord development. *Cell, 146,* 178.

Alvarez, F. J., & Fyffe, R. E. (2007). The continuing case for the Renshaw cell. *Journal of Physiology, 584,* 31–45.

Betley, J. N., Wright, C. V., Kawaguchi, Y., Erdélyi, F., Szabó, G., Jessell, T. M., & Kaltschmidt, J. A. (2009). Stringent specificity in the construction of a GABAergic presynaptic inhibitory circuit. *Cell, 139,* 161–174.

Boillee, S., Vande Velde, C., & Cleveland, D. W. (2006). ALS: A disease of motor neurons and their nonneuronal neighbors. *Neuron, 52,* 39–59.

Bonnot, A., Whelan, P. J., Mentis, G. Z., & O'Donovan, M. J. (2002). Locomotor-like activity generated by the neonatal mouse spinal cord. *Brain Research Reviews, 40,* 141–151.

Brown, T. G. (1911). The intrinsic factors in the act of progression in the mammal. *Proceedings of the Royal Society of London, 84,* 308–319.

Burke, R. E., Levine, D. N., Tsairis, P., & Zajac, F. E. (1973). Physiological types and histochemical profiles in motor units of the cat gastrocnemius. *Journal of Physiology (London), 234,* 723–748.

Butt, S. J., Lundfald, L., & Kiehn, O. (2005). EphA4 defines a class of excitatory locomotor-related interneurons. *Proceedings of the National Academy of Sciences of the United States of America, 102,* 14098–14103.

Cooper, E. C., & Jan, L. Y. (1999). Ion channel genes and human neurological diseases: Recent progress, prospects, and challenges. *Proceedings of the National Academy of Sciences of the USA, 96,* 4759–4766.

Cowley, K. C., & Schmidt, B. J. (1997). Regional distribution of the locomotor pattern-generating network in the neonatal rat spinal cord. *Journal of Neurophysiology, 77,* 247–259.

Crone, S. A., Quinlan, K. A., Zagoraiou, L., Droho, S., Restrepo, C. E., Lundfald, L., … Sharma, K. (2008). Genetic ablation of V2a ipsilateral interneurons disrupts left-right locomotor coordination in mammalian spinal cord. *Neuron, 60,* 70–83.

Drew, T., Prentice, S., & Schepens, B. (2004). Cortical and brainstem control of locomotion. *Progress in Brain Research, 143,* 251–261.

Eccles, J. C., Fatt, P., & Koketsu, K. (1954). Cholinergic and inhibitory synapses in a pathway from motor-axon collaterals to motoneurones. *Journal of Physiology, 126,* 524–562.

Gosgnach, S., Lanuza, G. M., Butt, S. J., Saueressig, H., Zhang, Y., Velasquez, T., … Goulding, M. (2006). V1 spinal neurons regulate the speed of vertebrate locomotor outputs. *Nature, 440,* 215–219.

Goulding, M. (2009). Circuits controlling vertebrate locomotion: moving in a new direction. *Nature Reviews. Neurosciences, 10,* 507–518.

Goulding, M., & Pfaff, S. L. (2005). Development of circuits that generate simple rhythmic behaviors in vertebrates. *Current Opinion in Neurobiology, 15,* 4–20.

Grillner, S. (1985). Neurobiological bases of rhythmic motor acts in vertebrates. *Science, 228,* 143–149.

Grillner, S. (2006). Biological pattern generation: the cellular and computational logic of networks in motion. *Neuron, 52,* 751–766.

Grillner, S., & Jessell, T. M. (2009). Measured motion: Searching for simplicity in spinal locomotor networks. *Current Opinion in Neurobiology, 19,* 572–586.

Henneman, E. (1957). Relation between size of neurons and their susceptibility to discharge. *Science, 126,* 1345–1347.

Hinckley, C. A., Hartley, R., Wu, L., Todd, A., & Ziskind-Conhaim, L. (2005). Locomotor-like rhythms in a genetically distinct cluster of interneurons in the mammalian spinal cord. *Journal of Neurophysiology, 93,* 1439–1449.

Jessell, T. M. (2000). Neuronal specification in the spinal cord: Inductive signals and transcriptional codes. *Nature Review Genetics, 1,* 20–29.

Kullander, K., Butt, S. J., Lebret, J. M., Lundfald, L., Restrepo, C. E., Rydstrom, A., … Kiehn, O. (2003). Role of EphA4 and EphrinB3 in local neuronal circuits that control walking. *Science, 299,* 1889–1892.

Lanuza, G. M., Gosgnach, S., Pierani, A., Jessell, T. M., & Goulding, M. (2004). Genetic identification of spinal interneurons that coordinate left-right locomotor activity necessary for walking movements. *Neuron, 42,* 375–386.

Lechner, H. A., Lein, E. S., & Callaway, E. M. (2002). A genetic method for selective and quickly reversible silencing of Mammalian neurons. *Journal of Neurosciences, 22,* 5287–5290.

Mentis, G. Z., Alvarez, F. J., Bonnot, A., Richards, D. S., Gonzalez-Forero, D., Zerda, R., & O'Donovan, M. J. (2005). Noncholinergic excitatory actions of motoneurons in the neonatal mammalian spinal cord. *Proceedings National Academy Sciences USA, 102,* 7344–7349.

Mentis, G. Z., Siembab, V. C., Zerda, R., O'Donovan, M. J., & Alvarez, F. J. (2006). Primary afferent synapses on developing and adult Renshaw cells. *Journal of Neurosciences, 26,* 13297–13310.

Nishimaru, H., Restrepo, C. E., Ryge, J., Yanagawa, Y., & Kiehn, O. (2005). Mammalian motor neurons corelease glutamate and acetylcholine at central synapses. *Proceedings of National Academy of Sciences USA, 102,* 5245–5249.

Petersen, N. T., Butler, J. E., Marchand-Pauvert, V., Fisher, R., Ledebt, A., Pyndt, H. S., … Nielsen, J. B. (2001). Suppression of EMG activity by transcranial magnetic stimulation in human subjects during walking. *Journal of Physiology (London), 537,* 651–656.

Petersen, N., Christensen, L. O., & Nielsen, J. (1998). The effect of transcranial magnetic stimulation on the soleus H reflex during human walking. *Journal of Physiology (London), 513,* 599–610.

Pratt, C. A., & Jordan, L. M. (1987). Ia inhibitory interneurons and Renshaw cells as contributors to the spinal mechanisms of fictive locomotion. *Journal of Neurophysiology, 57,* 56–71.

Renshaw, B. (1946). Central effects of centripetal impulses in axons of spinal central roots. *Journal of Neurophysiology, 9,* 191–204.

Rossignol, S., Drew, T., Brustein, E., & Jaing, W. (1999). Locomotor performance and adaptation after partial or complete spinal cord lesions in the cat. *Progress in Brain Research, 123,* 349–365.

Rudomin, P. (2009). In search of lost presynaptic inhibition. *Experimental Brain Research, 196,* 139–151.

Siembab, V. C., Smith, C. A., Zagoraiou, L., Berrocal, M. C., Mentis, G. Z., & Alvarez, F. J. (2010). Target selection of proprioceptive and motor axon synapses on neonatal V1-derived Ia inhibitory interneurons and Renshaw cells. *Journal of Comparative Neurology, 518,* 4675–4701.

Sinkjaer, T., Andersen, J. B., Ladouceur, M., Christensen, L. O., & Nielsen, J. B. (2000). Major role for sensory feedback in soleus EMG activity in the stance phase of walking in man. *Journal of Physiology (London), 523,* 817–827.

Tresch, M. C., & Kiehn, O. (1999). Coding of locomotor phase in populations of neurons in rostral and caudal segments of the neonatal rat lumbar spinal cord. *Journal of Neurophysiology, 82,* 3563–3574.

Wang, Z., Li, L., Goulding, M., & Frank, E. (2008). Early postnatal development of reciprocal Ia inhibition in the murine spinal cord. *Journal of Neurophysiology, 100,* 185–196.

Wilson, J. M., Hartley, R., Maxwell, D. J., Todd, A. J., Lieberam, I., Kaltschmidt, J. A., & Brownstone, R. M. (2005). Conditional rhythmicity of ventral spinal interneurons defined by expression of the Hb9 homeodomain protein. *Journal of Neurosciences, 25,* 5710–5719.

Zhang, Y., Narayan, S., Geiman, E., Lanuza, G. M., Velasquez, T., Shanks, B., … Goulding, M. (2008). V3 spinal neurons establish a robust and balanced locomotor rhythm during walking. *Neuron, 60,* 84–96.

Zhong, G., Droho, S., Crone, S. A., Dietz, S., Kwan, A. C., Webb, W. W., … Harris-Warrick, R. M. (2010). Electrophysiological characterization of V2a interneurons and their locomotor-related activity in the neonatal mouse spinal cord. *Journal of Neuroscience, 30,* 170–182.

Suggested Readings

Engel, A. G., & Sine, S. M. (2005). Current understanding of congenital myasthenic syndromes. *Current Opinion in Pharmacology, 5,* 308–321.

Kiehn, O. (2006). Locomotor circuits in the mammalian spinal cord. *Annual Review Neurosciences, 29,* 279–306.

O'Donovan, M. J., Bonnot, A., Wenner, P., & Mentis, G. Z. (2005). Calcium imaging of network function in the developing spinal cord. *Cell Calcium, 37,* 443–450.

Schwab, J. M., Brechtel, K., Mueller, C. A., Failli, V., Kaps, H. P., Tuli, S. K., & Schluesener, H. J. (2006). Experimental strategies to promote spinal cord regeneration—an integrative perspective. *Progress in Neurobiology, 78,* 91–116.

George Z. Mentis

CHAPTER

29

Control of Movement

Four major pathways descend from the brain to the spinal cord to control muscles that move the skeleton. These pathways arise in the vestibular nuclei (vestibulospinal), in the brainstem reticular formation (reticulospinal), in the red nucleus (rubrospinal), and in the cerebral cortex (corticospinal). While these four descending pathways work together to provide seamless control of movements from postural reflexes to delicate manipulation, they can be separated broadly into two main systems. Vestibulospinal and reticulospinal pathways can be grouped together as a medial system: their axons descend through the brainstem and spinal cord close to the midline, and chiefly innervate axial and proximal limb musculature in which motoneurons lie medially in the ventral horn. The corticospinal and rubrospinal pathways together can be considered a lateral system: their axons descend in the lateral column of the spinal cord and chiefly innervate limb musculature, particularly distal musculature, in which motoneurons lie laterally in the ventral horn. More than just an anatomical distinction, the medial and lateral systems have different principal functions. The medial system provides postural control. Monkeys with medial system lesions fall over when they attempt to ambulate and climb, but when supported, they can use their hands and fingers adeptly in retrieving food pieces from narrow holes (Lawrence & Kuypers, 1968b). The lateral system provides fine control of voluntary movement. Monkeys with lateral system lesions rapidly recover the use of all four extremities in activities such as ambulation and climbing, but they are unable to make the fine finger movements needed to extract small pieces of food from narrow holes (Lawrence & Kuypers, 1968a). This chapter examines the contributions of first the medial and then the lateral systems to control of movement.

THE MEDIAL POSTURAL SYSTEM

Vestibular and Reticular Nuclei Control Posture and Reflex Behaviors

Early experimenters found that some degree of postural and reflex control remains after higher brain centers are completely cut off from the brainstem and spinal cord, and that the extent of preserved motor function is determined by the level of the transection that disconnects the upper portions of the brain. There is no functional regeneration of central neuronal pathways after damage, and when the spinal cord is separated from all of the brain above, as can occur in motor vehicle or equestrian accidents, the result is quadriplegia, loss of all voluntary movement and postural control in the arms and legs. Only the spinal reflexes discussed in the preceding chapter return, and those return only when spinal cord neurons recover excitability after a prolonged period of limp (flaccid) paralysis called spinal shock.

Transections at higher levels of the neuraxis leave part of the brainstem connected with the spinal cord. Chronic bulbospinal cats, those with medulla and spinal cord below the transection, show primitive attempts at righting themselves when placed on one side. Effective righting may be observed in chronically surviving animals subjected to transection that leaves the mesencephalon intact below the level of the cut. As the level of transection is raised, posture and balance become progressively closer to normal. The decorticate animal produced by the ablation of cerebral cortical tissue has many apparently normal postural reactions, lacking only certain placing and stepping reactions. Progressively higher brain transections enable more of the vestibular (balance) and proprioceptive (muscle

© 2013 Elsevier Inc. All rights reserved.

and position sense) circuitry that participates in postural control, as well as improving the overall balance between excitatory and inhibitory influences on the spinal cord.

Nuclei of the brainstem reticular formation with intact spinal projections after high brainstem transaction include the locus ceruleus and raphe nuclei, the origins of descending noradrenergic and serotonergic neuromodulatory axonal projections that alter excitability of neurons in the spinal cord, as discussed in the preceding chapter. Parts of the raphe nuclei also have a role in circadian rhythms (Chapter 39), and projections of the locus ceruleus are involved in attention (Chapter 46) and internal reward (Chapter 41). Other areas of the reticular formation involved in motor control have specific functions in generation of the locomotor rhythm (MLR, mesencephalic locomotor region, Chapter 27) and rapid eye movements (PPRF and MRF, Chapter 32). Nuclei of the reticular formation are most often studied from the perspective of their contributions to specific functional systems described in other chapters and through the actions of their monoamine neuromodulators noradrenalin (norepinephrine) and serotonin. Vestibular nuclei influence several brain regions and are studied for their contributions to spatial cognition, autonomic function, eye movements, and postural reflexes.

The Vestibular Apparatus Senses Head Rotation and Tilt

The sensory inputs most important for postural responses are visual, proprioceptive, and vestibular. Vestibular sensory signals from the labyrinth of the inner ear provide direct information about rotations of the head and its orientation with respect to gravity. Two types of vestibular transducer organs—semicircular canals and otolithic maculae—are located within the labyrinth of the ear and respond to accelerations of the head in space. The three semicircular canals and the two otolithic maculae are stimulated by rotational and linear acceleration, respectively. The mechanical structure of the sense organ determines the nature of the effective stimulus (Wilson & Melvill-Jones, 1979).

The labyrinth (Fig. 29.1) is outlined by the bony labyrinth, a set of passages in the skull. The bony labyrinth is lined with membranes that contain endolymph and which make up the membranous labyrinth. The membranous labyrinth has a large central vessel called the utricle, another vessel called the saccule, and three narrow passages that emerge from the utricle and loop around to rejoin it. These three loops form circular passages for endolymph and are known as semicircular canals. The canals are oriented orthogonal to one another: there is a lateral or horizontal canal, an anterior or superior canal, and a posterior canal. A semicircular canal is excited by rotational accelerations, and the mechanical structure of the canal converts acceleration into a rotational velocity signal. The three canals in each labyrinth are oriented for excitation by rotations toward that side of the body, horizontally, diagonally forward, and diagonally backward for the horizontal, anterior, and posterior canals, respectively. Thus, each labyrinth is excited by head rotation toward its side, called ipsilateral rotation. Canal afferents fire action potentials at a rate approximately proportional to the velocity of head rotation in the direction of the canal orientation. In contrast with the auditory system, in which 8th nerve afferents are narrowly tuned to specific frequencies of sound (see Chapter 25), vestibular nerve afferents are broadly tuned, responding to the entire range of head motion frequencies.

There are two otolith organs in mammals: the utricle (utriculus) and saccule (sacculus). The names apply to the fluid chambers in which they lie, and the specific locations of the receptors are called the maculae. The two maculae are small regions that contain hair cells innervated by vestibular afferents. Like auditory and semicircular canal hair cells, bending of their cilia is the excitatory stimulus for macular hair cells, and the cilia are imbedded in a gelatinous mass above the cell bodies. Unlike canal hair cells, the otolithic hair cell maculae rest in large chambers, in which fluid motion is thought to exert little or no force on the cilia. Instead, the gelatinous mass holding the cilia contains dense crystals, the otoconia or statoconia, and the entire mass above the hair cells of a macula is acted on by gravity or linear acceleration because of its higher density than the endolymph. The otoconial mass sags in the direction of a head tilt. This sagging, or the equivalent lagging of the mass behind a linear acceleration, provides the natural stimulus for otoliths.

The utricular macula lies horizontally on the floor of its cavity. When the head is still and held approximately erect with respect to gravity, there is no stimulus to this organ, but tilts from the erect posture will excite utricular afferents. Microscopic examination of the utricular macula shows that the hair cells are not all oriented with their cilia arranged in the same direction, and different directions of tilt excite different populations of utricular afferents. The direction of the hair cell, indicated by its structural orientation, corresponds to the best direction of tilt for exciting the receptor. Each utricular macula has hair cell orientations capable of responding to any direction of head tilt, but ipsilateral tilt excites a predominant number of afferents. Thus, as for the semicircular canals, we say that ipsilaterally directed head motion is excitatory. The saccular macula is placed vertically on the side of the saccule, approximately in a parasagittal plane. There is a strong acceleration stimulus to the saccular macula when the head is erect and little or no stimulus when the head is lying on either side. The sacculus responds vigorously when the head bobs up and down, as during locomotion. In summary, two types of vestibular signals are available to assist balance: ipsilaterally directed rotational head

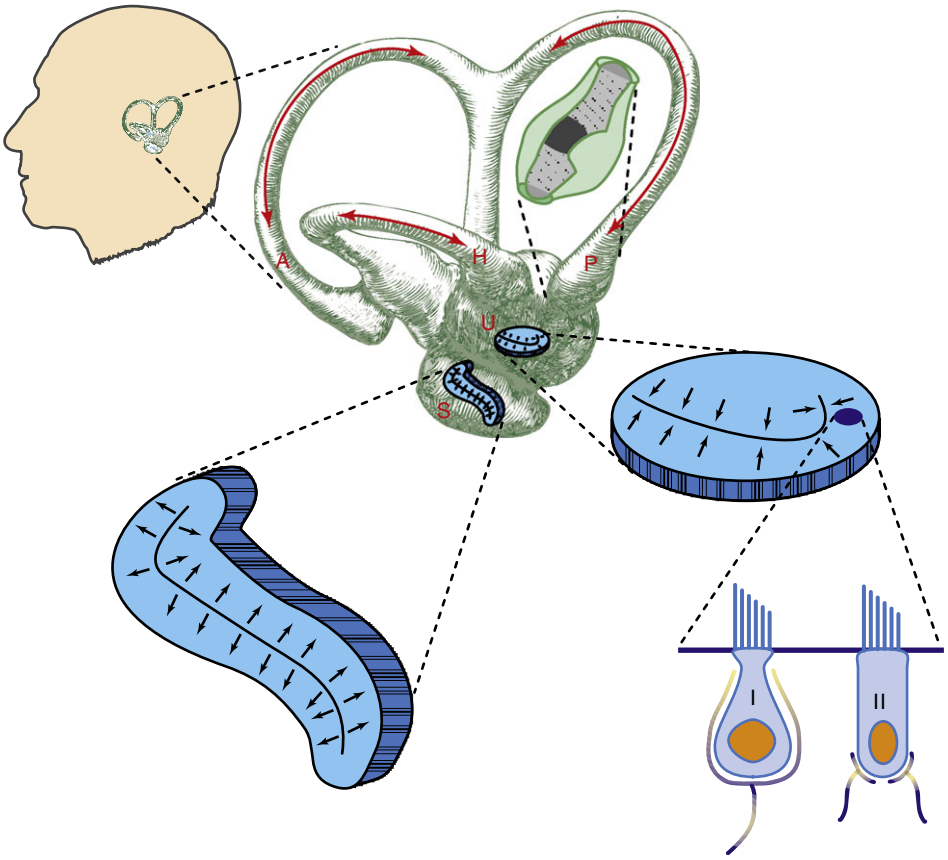


FIGURE 29.1 Vestibular canals and otoliths. The position and orientation of the labyrinth (not to scale) in the head are shown at the upper left of the figure. An enlarged view of the labyrinth shows the directions of head rotation and endolymph flow (red arrows) that excite each of the three semicircular canals. The horizontal orientation of the utricular macula and vertical orientation of the saccular macula are shown schematically. More highly magnified views of the receptor regions of a canal and of the otolith organs are shown with the cupula of the canal colored dark gray and the best directions for excitation of otolith hair cells marked by black arrows. At the lower right are two anatomical types of hair cell: the calyx or type I and bouton-ending or type II receptor. The tallest cilial extension on each cell is the kinocilium. A, anterior semicircular canal; H, horizontal semicircular canal; P, posterior semicircular canal; S, saccule containing saccular macula; U, utricule containing utricular macula; I, type I receptor; II, type II receptor. *Based on studies reviewed by Wilson and Melvill-Jones (1979).*

velocity is signaled by the semicircular canals and ipsilaterally directed head tilt or an equivalent linear acceleration is signaled by the utricular otoliths.

Vestibular Nuclei Convey Head Motion Signals to Many Brain Regions

Four brainstem vestibular nuclei receive inputs from the vestibular portion of the VIIIth nerve (Fig. 29.2). The medial, lateral (Deiter's), and descending (inferior) nuclei contribute to the medial and lateral vestibulospinal tracts which provide direct vestibular influence on the spinal cord. Neurons in the superior and medial vestibular nuclei mediate vestibulo-ocular reflexes (Chapter 32). The medial vestibular nucleus also sends projections to brainstem autonomic nuclei and to the parabrachial nucleus, which integrates viscerosensory signals and has bidirectional connections with the amygdala. Ascending pathways from each of the vestibular nuclei project to thalamic nuclei, which influence widespread regions of the cortex.

Neurons in the Vestibular System Are Specialized for Fast Behaviors

To effectively stabilize the body and eyes during self-motion, the vestibular system must rapidly convey head motion signals to motoneurons. The vestibular system is the fastest of all sensory systems; rapid head movements can evoke eye movements in less than 10 msec. Several mechanisms ensure that the central processing of head motion signals is both fast and accurate. Vestibular nerve afferents and central vestibular neurons fire tonically at high rates—about 50–100 Hz when the head is not moving—and modulate their firing rates in response to head movement. High firing rates enable head motion signals to be encoded both rapidly and bidirectionally; for example, vestibular nucleus neurons receiving inputs from horizontal canal afferents increase firing rates when the head turns to the same side (ipsiversively) and decrease firing rates when the head turns to the opposite side (contraversively). As with vestibular afferents, central vestibular nucleus neurons are broadly tuned for head movement frequencies and their firing rates are linearly modulated by head movement.

Central vestibular synapses and neurons are specialized for linear signaling over a wide range of firing rates. At well-studied synapses onto pyramidal cells in the hippocampus and cortex, synaptic efficacy is influenced strongly by the recent history of presynaptic activity (see Chapter 10). In contrast, across the behaviorally relevant range of afferent firing rates, central vestibular nerve synapses exhibit neither short-term depression nor

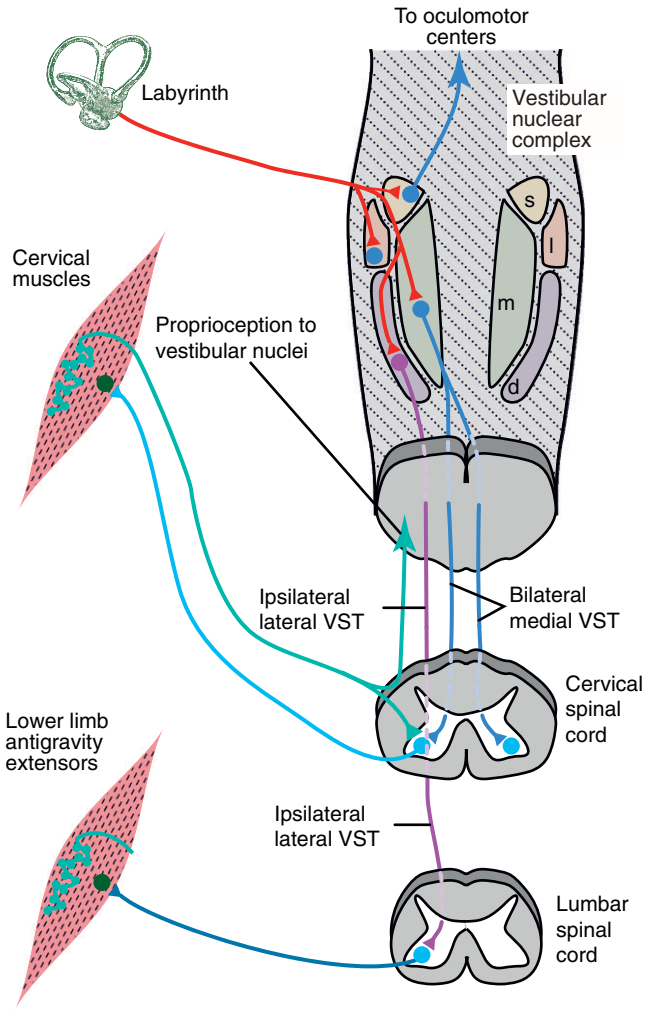


FIGURE 29.2 Vestibular and proprioceptive reflex signal inputs and major pathways from the brainstem vestibular nuclei. The medial vestibulospinal tract projects bilaterally to the cervical spinal cord to mediate the vestibulocollic reflex. The lateral vestibulospinal tract descends to lumbar levels of the spinal cord to influence limb extensors involved in balance. Neck muscle proprioceptors send signals to vestibular nuclei to participate in cervicocollic reflexes and interactions among reflexes. d, descending vestibular nucleus; 1, lateral vestibular nucleus; m, medial vestibular nucleus; s, superior vestibular nucleus; VST, vestibulospinal tract.

short-term facilitation, enabling them to transmit signals that accurately reflect head motion (Bagnall, McElvain, Faulstich, & du Lac, 2008). Postsynaptic vestibular nucleus neurons fire spontaneously in the absence of synaptic stimulation, enabling rapid responses to synaptic inputs. They express sodium and potassium channels that enable them to sustain firing at high rates (see Chapter 31) as well as calcium-activated potassium channels which enable them to respond to synaptic inputs with linear changes in firing rates (Kolkman, McElvain, & du Lac, 2011). These synaptic and cell biological properties of vestibular neurons enable vestibulo-motor reflexes to respond rapidly and linearly to head motion.

Vestibulocervical and Vestibulospinal Reflexes Stabilize Head and Body Posture

Vestibulocervical (neck or collic reflexes) and vestibulospinal reflexes use canal and otolith signals to stabilize the posture of the head and body. These reflexes act as "negative feedback" systems (Schor, Kearney, & Dieringer, 1988). When the head and body rotate or tilt in any direction, the vestibular stimulus excites pathways that contract neck and limb muscles that oppose the motion so that the undesired movement is reduced or corrected. The importance of vestibulocervical and vestibulospinal reflexes is demonstrated by damage to the labyrinth or VIIIth cranial nerve. Unilateral labyrinthectomy causes an initial postural disability, with leaning or falling toward the side of the lesion. Bilaterally symmetric semicircular canal plugging, which removes head velocity signals with little loss or imbalance in tonic vestibular nerve activity, produces head instability with oscillations that may persist for several days.

Electrical activation of the vestibular pathways shows that vestibulocervical reflexes are mediated primarily by excitatory and inhibitory connections of the bilaterally projecting medial vestibulospinal tract and that ipsilateral excitation of leg antigravity muscles is carried via the lateral vestibulospinal tract. The bulk of the lateral vestibulospinal tract projects from the lateral vestibular nucleus to the lumbar spinal cord for vestibulospinal reflex control of upright body posture. Vestibular nuclei also project heavily to reticular nuclei, and only lesions that impinge on both vestibulospinal and reticulospinal pathways substantially blunt vestibular reflex responses.

The bilateral, mixed excitatory and inhibitory connections of the vestibulocervical reflex circuits indicate that its actions are complex, and considerable effort has been devoted to explicating the exact nature of these actions. The excitatory lateral vestibulospinal tract projection to ipsilateral extensor motoneurons offers a simple example of negative feedback: When the head and body tilt to one side, canals and otoliths of the labyrinth on that side are excited, and the lateral vestibulospinal tract carries this excitation to the ipsilateral leg extensor motoneurons. The ipsilateral leg is thus extended to oppose the tilt and maintain upright posture.

Initial quantitative studies of the vestibulocervical reflex used sinusoidally oscillating stimuli and concentrated on the timing of the reflex head torques or the electromyographic activity that accompanied neck muscle contractions The conclusion from these studies was that central vestibular circuits process low frequency (slow) head movements and high frequency (fast) movements in distinct ways. At low frequencies a portion of the velocity signal from the semicircular canals is converted to a head position signal appropriate for repositioning the head in response to a slow change in head angle. At high

frequencies, however, vestibulocervical reflex circuitry introduces a "phase lead," or anticipatory response, which works in conjunction with the canal afferent signal to produce movements with short response times. As such, the reflexive behavioral responses oppose head angular acceleration, as would be appropriate given that inertia of the head must be overcome for adequate compensation during rapid head motion.

The limbs and trunk represent a more complex mechanical system than the neck and head, complicating the analysis of vestibulo-spinal reflexes. Electromyographic recordings in decerebrated animals have shown that limb extensors at low frequencies of vestibular oscillation are excited in phase with head position. For example, when the head rolls slowly to the left, the left forelimb extends as if to brace against further displacement. The likely source of signals exciting these responses is the otolith organs. As the frequency of head oscillation increases, the timing of vestibulospinal responses corresponds more closely to head rotation velocity, indicating that semicircular canal signals begin to predominate over otolith signals in vestibulospinal responses at higher frequencies.

Body musculature could be coordinated to compensate for a particular direction of head and body rotation using a variety of strategies, but only general features of vestibulocervical and vestibulospinal synergies are known (Baker, Goldberg, & Peterson, 1985). Each of the major neck muscles responds to vestibular stimulation (rotation) in many different directions, specific for each neck muscle, that do not reflect the directional sensitivity of any single semicircular canal. Instead, the responses to rapid motions reflect a weighted sum of canal and otolith inputs. The appropriate postural response to a particular pattern of vestibular stimulation depends on many factors, including the position of the head, trunk, and limbs. The cerebellum is critical for the complex, context-dependent computations that mediate vestibulo-spinal reflexes (Chapter 31).

The Cervicocervical Reflex Stabilizes the Head by Opposing Lengthening of Neck Muscles

The proprioceptive contribution to reflex stabilization of the head and body has been examined in many of the same ways as vestibular contributions (Peterson, Goldberg, Bilotto, & Fuller, 1985). The use of proprioceptive signals is more complicated than for vestibular signals, as any of the muscles of the neck, trunk, or limbs could provide an input to influence any other muscle or synergistic group of muscles. The neck stretch or cervicocervical reflex has been studied most extensively. Like the stretch reflex for limb extensors, the cervicocervical reflex opposes lengthening of the muscles concerned and so is a negative feedback compensatory system.

The cervicocervical reflex has been analyzed by rotating the trunk of experimental animals while holding the head fixed in space. This stimulates neck proprioceptors without stimulating vestibular sense organs. Nevertheless, the results are reminiscent of the vestibulocervical reflex. At low frequencies of neck rotation cervicocervical reflex muscle contractions correspond to the extent of muscle stretch, but at higher frequencies the responses occur more rapidly, corresponding to the velocity or acceleration of muscle stretch. The directionality of neck muscle responses to stretch matches that of vestibulocervical responses fairly well. Common directionality implies that signals for vestibular and stretch reflexes to neck muscles share central circuitry.

The Brainstem Controls Coordinated Postural Reactions

A variety of reflexes contribute to the maintenance of overall body posture. The tonic neck or cervicospinal reflexes are the best known of these, but the supporting reactions, placing reactions, righting reactions, hopping or stepping reactions, and others are also useful parts of the neural control of posture. The tonic neck reflex adjusts the extension of the limbs in response to the angle of the head on the trunk. Roberts (1967) showed that stretching the neck by rolling the spine to one side elicited tonic flexion and extension of the forelimbs, as if the support surface had been tilted to produce the rotation. The placement of a limb on the ground initiates a set of reflex reactions that stiffen the limb into a supporting pillar. This response is called the "positive supporting reaction" and depends on the integrity of the brainstem. As discussed earlier, righting reactions also depend on the brainstem, and "optical righting reflexes" mediated by vision require an intact cerebral cortex. Stable posture requires that the feet be not only rigid but placed correctly on a supporting surface, and the placing reactions of quadrupeds contribute to this by moving the feet toward a visible surface (visual placing reaction) or onto a surface that has tactile contact with the top of the foot, chin, or whiskers (tactile placing reactions). Although a tactile placing reaction may be evoked in primitive form in spinalized animals, it is generally held that at least the lower brainstem must be intact for tactile placing to occur effectively. Finally, if balance reactions are insufficient to maintain stable posture, as when the support surface is moved beneath the foot, the limb may hop to a new position where stable posture is possible (hopping reaction). The action of the righting reflex of cats during falling is familiar to all. Humans show a similar (but less graceful) reaction to an unexpected drop, measurable as a short latency electromyographic response in the gastrocnemius muscle. In cats, this response survives blockage of the semicircular canals, but not total labyrinthectomy, and so appears to be mediated by otoliths.

Balance Depends on Context-Dependent Postural Strategies

An important tool for exploring body stability is the posture platform, a base on which subjects can stand and be subjected to displacements of the underlying supporting surface. The posture platform is used with body position sensors and electromyography to evaluate standing posture (Horak & Nashner, 1986; Horak & Shupert, 1994; Nashner, 1982). Even on a perfectly stable support surface with vision, somatosensory, and vestibular senses all active, a standing person will sway slightly. When the eyes are closed, sway is greater. When the visual surround is linked to the subject's head position, visual information is misleading and sway is increased. When the support surface tilts along with the body, sway increases because proprioceptive information is misleading. The vestibular system thus works in combination with proprioception and vision to promote postural stability. How multisensory signals are appropriately weighted and combined to produce a reliable and integrated sense of self motion is an area of active research (Angelaki & Cullen, 2008)

Sudden displacement of the platform on which a subject stands allows measurement of the latencies and patterns of electromyographic activity in the postural muscles of the legs. When the subject is standing on a large platform that is displaced forward or backward, the leg muscles contract in sequence from ankle to thigh to hip, and motion occurs primarily about the ankle joint. When the platform displacement is forward, the ankle flexor tibialis anterior on the front of the calf is first to be excited (flexion is the toe-up direction), at about 80–100 ms after the displacement. This is followed about 20 ms later by contraction of the quadriceps thigh muscles and then later still by contraction of trunk musculature. Backward displacement first elicits excitation of the gactrocnemius muscle at the back of the calf to extend the ankle, followed by excitation of thigh and trunk muscles antagonistic to those excited by forward displacement (Fig. 29.3). The overall postural reaction of distal-to-proximal excitation has been termed an "ankle strategy" for maintaining balance.

Although the ankle strategy dominates in simple forward–backward displacement, other strategies are available, and postural reactions adapt to alterations in the support surface or sensory inputs. Displacement of a support surface that is short compared to the foot, or rotation of the support instead of displacement, demands a different postural response. In these cases a hip strategy is adopted, in which the body is bent at the hips so that the lower half of the body moves in the same direction as it does during ankle strategy reactions while the upper body moves in the opposite direction. Thus, during a forward, front downward tilt of a supporting platform, the hips move backward and the head forward. This shifts the center of gravity, which has moved forward due to the displacement,

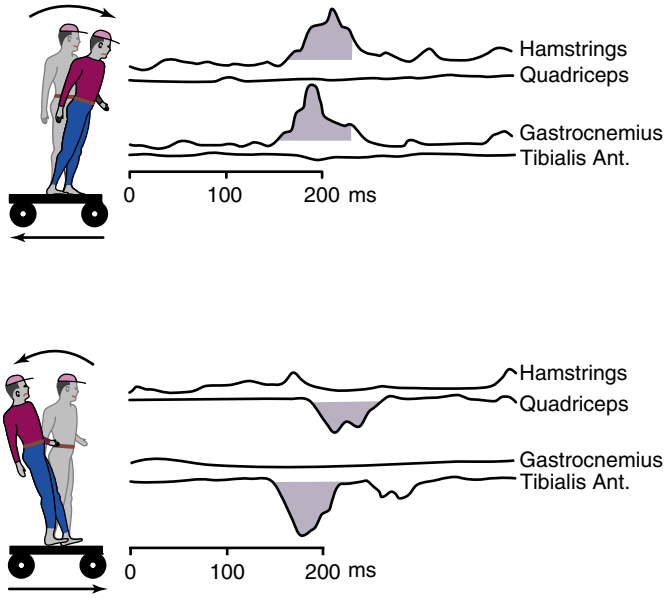


FIGURE 29.3 Sequencing of muscle activation in response to displacement of a supporting platform. When the supporting surface is displaced backward (at 0 ms), flexor muscles are excited first in the distal lower limb segments (gastrocnemius, about 80 ms latency) and then in the proximal segment (hamstrings, about 100 ms latency). Forward displacement of the platform activates lower limb extensors, again in a distal (tibialis anterior) to proximal (quadriceps) sequence. Black arrows mark the first detected electromyographic response to displacement. *Based on studies by Horak and Nashner (1986).*

backward so that it is placed over the support again. Other work has documented a stiffening strategy of muscle cocontraction in response to base rotation and a multilink strategy that includes ankle, hip, and neck joint motions in response to translation. Stiffening, ankle, and hip or multilink strategies are applicable to the sway that occurs during normal standing on commonly encountered kinds of surfaces and may represent basic units of postural reaction. The range of postural reactions recorded from standing human subjects argues for a complexity beyond that observed in the vestibulocervical or vestibulospinal reflexes studied during rotation or tilting of animal subjects.

Vestibular Damage Results in Disorders of Postural Control

Control of posture involves many regions of the nervous system, and it is not surprising that disorders of posture can result from damage to the sensory periphery or to telencephalic, cerebellar, brainstem, or spinal centers. The striking motor consequences of lesions of the basal ganglia may include profound effects on posture, such as the rigidity and general poverty of movement associated with Parkinson's disease (Chapter 30). Damage to the anterior vermis of the cerebellum can exaggerate decerebrate rigidity, and cerebellar patients show poorer performance

BOX 29.1

VESTIBULAR PLASTICITY

The experience of sailors getting their "sea legs" demonstrates that postural reflexes and balance can adapt over time to produce better performance in new circumstances. Clear evidence for this comes from the study of postural control in astronauts. After a space flight, astronauts rely more heavily than before on visual cues for postural orientation, and they show degraded performance in their responses to disturbances generated by a posture platform. Sway during standing is increased dramatically when visual cues are removed, and the body segments move in a less coordinated manner than before space flight. Nonetheless, the performance of astronauts after the experience of space flight is quite remarkable; they are able to balance adequately within hours of landing and regain their preflight postural performance within a few days.

The behavioral strategies and cellular mechanisms that mediate compensation for vestibular loss have been studied in animal models. Following unilateral labyrinthectomy (surgical removal of the inner ear), animals exhibit transient postural impairments, such as rolling and tilting toward the side of the lesion, which resolve within hours. Two distinct mechanisms account for the rapid recovery of postural stability despite the persistent impairment of vestibular function. First, the influence of the remaining senses that signal self-motion—proprioception and vision—increases rapidly following vestibular loss. Second, cellular mechanisms of plasticity in vestibular nucleus neurons enable the available sensory signals to recalibrate postural reflexes.

James F. Baker, Marc H. Schieber and Sascha du Lac

Reference

Cullen, K. E., Minor, L. B., Beraneck, M., & Sadeghi, S. G. (2009). Neural substrates underlying vstibular compensation: Contributions of peripheral versus central processing. *Journal of Vestibular Research, 19,* 171–182.

in posture platform situations, with less adaptive modification of responses, as expected from our knowledge of cerebellar function (Chapter 31).

Many types of disease or damage of the vestibular system affect posture (and generally also cause dizziness), including vestibular neuritis, peripheral or central tumors or infarction, and Meniere's syndrome (Baloh & Honrubia, 1990) (see Box 29.2). Bilateral involvement and acute, episodic, or chronic time courses occur in many vestibular disorders. The most obvious form of vestibular system damage is unilateral labyrinthectomy, performed either experimentally in animals or as the result of disease or surgical intervention against disease in humans. The postural symptoms immediately following loss of vestibular input to one side of the brain vary across species, with nonmammalian and "lower" mammalian species typically showing a more prominent tilting of the head and body toward the side of damage. Vestibular afferents have a high level of resting activity, and leaning toward the side of lost vestibular signals makes sense in that the spontaneous or resting activity of the remaining vestibular apparatus is signaling, in effect, motion toward the intact side because it is not balanced by equal resting activity from the damaged side. To counter this neural stimulus, the affected individual tilts away from the directions of apparent motion. This static postural effect of unilateral labyrinthectomy is accompanied by dynamic postural deficits, seen in experimental animals as weakened ipsilateral limb extensor responses and delayed responses to sudden drops. Remarkable compensation for unilateral labyrinthectomy occurs over a period of a few weeks, during which there is considerable recovery (see Box 29.1).

Patients with long-standing bilateral vestibular loss may perform well on posture platforms when visual and somatosensory cues are present, but fail completely to maintain upright stance when the support surface and visual surround both sway with the patient so that only vestibular information is accurate. Patients with recent bilateral vestibular loss or who have not yet compensated for vestibular loss also do poorly when vision and support surface are unreliable, but these patients perform poorly if only one sense, be it vision or somatosensory, is made an unreliable indicator of balance.

Summary

Axons from vestibular and reticular nuclei in the brainstem descend in pathways located ventromedially in the spinal cord to provide postural tone and balance. Balance is maintained by negative feedback reflexes that are stimulated by the vestibular canals and otoliths. The vestibulocervical reflex excites neck muscles via the medial vestibulospinal tract to oppose head tilt, and vestibulospinal reflexes carried by the lateral vestibulospinal tract excite ipsilateral limb extensors to oppose body tilt. Vestibular, muscle stretch, and other reflexes work together to coordinate posture and balance, and the medial system adopts different strategies for balance control depending

BOX 29.2

MENIERE SYNDROME

The typical patient with Meniere syndrome develops a sensation of fullness and pressure along with decreased hearing and tinnitus in one ear. Vertigo follows rapidly, reaching a maximum intensity within minutes and then slowly subsiding over several hours. Often the patient is left with a sense of unsteadiness and nonspecific dizziness that can go on for days after the acute vertiginous spell. In the early stages, hearing loss is completely reversible, but as the disease progresses, residual hearing loss becomes a prominent feature. The tinnitus is typically described as a roaring sound similar to the sound of the ocean. These episodes occur at irregular intervals over years, with periods of remission unpredictably intermixed. Eventually, most patients reach the so-called "burnt-out phase," where the episodic vertigo disappears and severe permanent hearing loss remains.

The clinical syndrome was first described by Prosper Meniere in 1861, but Hallpike and Cairns made the initial clinical-pathological correlation with hydrops of the labyrinth in 1938. Patients with Meniere syndrome invariably show an increase in volume of endolymph associated with distention of the entire endolymphatic system. Herniations and ruptures in the membranous labyrinth commonly occur, which may explain the episodes of hearing loss and vertigo.

Delayed endolymphatic hydrops occurs in an ear that has been damaged years before, usually by infection. With this disorder, the patient reports a long history of hearing loss, typically since childhood, followed many years later by episodic vertigo but without the typical auditory symptoms. The pathologic findings are remarkably similar to idiopathic Meniere syndrome, suggesting a common etiology. A subclinical viral infection could damage the resorptive mechanism of the inner ear, leading to an eventual decompensation in the balance between secretion and resorption of endolymph.

The key to the diagnosis of Meniere syndrome is to document fluctuating hearing levels in a patient with the characteristic clinical history. In the early stages, the sensorineural hearing loss is usually greater in the low frequencies. Some patients with Meniere syndrome develop abrupt episodes of falling to the ground without loss of consciousness or associated neurologic symptoms. These episodes have been called "otolithic catastrophes" because they are thought to result from a sudden mechanical deformation of the otolith receptor organ. Patients often report feeling as though they were pushed to the ground by some external force.

Because the cause of Meniere syndrome is usually unknown, treatment is empiric. Medical management consists of symptomatic treatment with antivertiginous drugs and long-term prophylaxis with salt restriction and diuretics. Many different surgical procedures have been tried but none has been consistently effective. Shunt operations to decrease the endolymph pressure have not been successful because the implanted drain devices are encapsulated rapidly by fibrous tissue. Destructive surgeries (removing the labyrinth or cutting the vestibular nerve), can stop the episodes of vertigo but do not change the tinnitus and progressive hearing loss.

Robert W. Baloh

on the body support surface and information from sensory inputs. The medial postural system can adapt to different postural situations, but its actions can be compromised by central neural damage or unbalanced by peripheral vestibular loss. The reflexive control of body stability and posture by the medial descending system allows the lateral descending system to specialize in execution of precise voluntary movements of the extremities.

THE LATERAL VOLUNTARY SYSTEM

While the medial system thus provides underlying postural control for stance, ambulation, and orientation of the head, the lateral system superimposes the ability to make more sophisticated, voluntary movements in response to complex features of the external environment (perceived through the senses) and internal state (stored memories, knowledge, and emotion). Much of this control is mediated by specialized regions of the cerebral cortex. Outflow from this motor cortex dominates the lateral descending system, particularly in primates and humans.

Components of the Lateral Voluntary System

The Corticospinal Projection Is the Most Direct Pathway from the Cerebral Cortex to Spinal Motoneurons

For more than a century, electrical stimulation of a limited portion of the frontal lobe of the cerebral cortex has been known to evoke movements. Electrical stimulation of this "excitable cortex" evokes movements readily because this region has relatively direct connections to

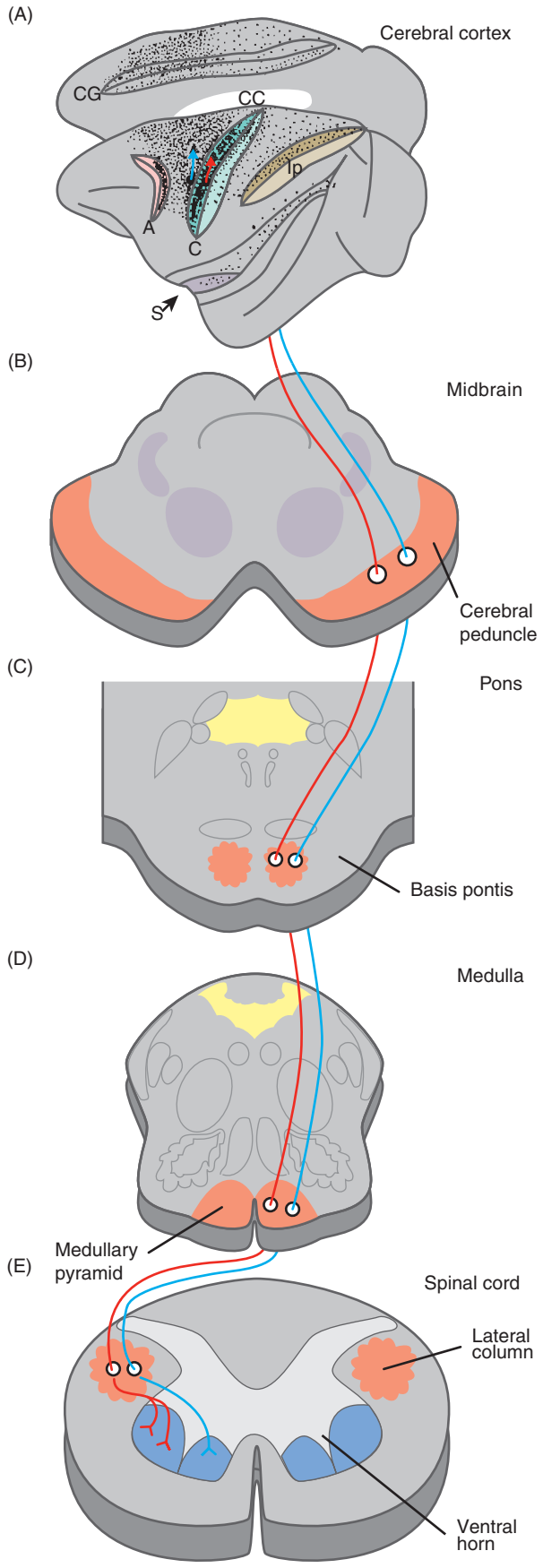


FIGURE 29.4 The corticospinal projection in the macaque monkey. (A) The density of corticospinal neuronal somata is shown by stippling in this lateral view of the left cerebral hemisphere; the superior medial surface of the hemisphere is also shown (above) as if reflected in a mirror. The central sulcus (C), arcuate sulcus (A), cingulate sulcus (Cg), intraparietal sulcus (Ip), and Sylvian fissure (S) are drawn as if pulled open to reveal the neurons in their banks. Two schematic corticospinal neurons, one relatively posterior and the other relatively anterior in area 4, send their axons down through the midbrain (B), pons (C), medulla (D), and spinal cord (E), which are drawn in cross section. In the spinal cord, the former corticospinal axon leaves the lateral column to terminate in the dorsolateral ventral horn, whereas the latter axon terminates in the ventromedial ventral horn.

spinal motoneurons. In primates with a central sulcus (Rolandic fissure) in the neocortex, cortical neurons with axons projecting to the spinal cord are found most densely in the anterior bank of the central sulcus. The density of such neurons decreases from there rostrally to the precentral sulcus (posterior bank of the arcuate sulcus in macaque monkeys) and medially to the cingulate sulcus (Fig. 29.4). This territory corresponds to cytoarchitectonic areas 4 and 6 of Brodmann. Additional corticospinal neurons in areas 1, 2, 3, 5, and 7 of the parietal lobe project to the dorsal horn of the spinal cord to regulate sensory inflow.

The axons of neurons that project from the cerebral cortex to the brainstem and spinal cord have large pyramid-shaped somata in cortical layer V. Their axons leave the cortex, pass through the centrum semiovale, and enter the internal capsule, along with many other axons from other cortical areas. Axons from the motor cortex are concentrated most heavily in the middle third of the posterior limb of the internal capsule. As axons descend from the internal capsule below the thalamus, they come to lie on the ventral surface of the brainstem. Here, they form the cerebral peduncle of the midbrain, where the corticospinal axons are concentrated in the middle third. Some descending axons end in or send collaterals to the red nucleus. At the level of the pons, the descending axons of the peduncle become intermixed with pontine nuclear neurons in the base of the pons (basis pontis). Most of these axons synapse on pontine neurons, providing input from the cerebral cortex to the cerebellum. Other descending axons synapse in the pontomedullary reticular formation, providing input to the reticulospinal system and to brainstem motor pattern generators (Chapter 27). The continuing corticospinal axons collect to form the medullary pyramid (Box 29.3). As the medulla blends into the spinal cord, the vast majority of corticospinal axons cross the midline and enter the lateral column of white matter on the opposite side of the spinal cord. Because of this decussation of the pyramidal tract, the left motor cortex controls movements on the right side of the body and vice versa. A small minority of corticospinal axons

BOX 29.3

PYRAMIDS IN THE BRAIN

Inspection of the ventral surface of the medulla reveals a structure that looks like a small pyramid. This medullary pyramid turns out to be the collected bundle of corticospinal axons. The corticospinal tract therefore is often referred to as the pyramidal tract. Axons of noncorticospinal descending pathways do not pass through the medullary pyramid and therefore can be described collectively as extrapyramidal pathways. At one time these pathways were thought to receive their major controlling inputs from the basal ganglia. Human neurologic disorders that result from dysfunction of the basal ganglia thus came to be known as "extrapyramidal syndromes." This term is misleading, however. More recent work has shown that many of the brainstem nuclei giving rise to extrapyramidal pathways receive cortical input and that much of the output of the basal ganglia is directed via the thalamus back to cortical motor areas. Many manifestations of extrapyramidal syndromes may therefore be played out through cortical motor areas, even via the pyramidal tract. Indeed, surgical lesions of the pyramidal tract once were used to ameliorate some extrapyramidal syndromes.

Marc H. Schieber and James F. Baker

remain uncrossed and continue caudally as the ventral (or anterior) corticospinal tract near the ventral midline of the spinal cord.

As descending corticospinal axons reach their target levels in the spinal cord, they enter the spinal gray matter, where they ramify and synapse. The majority of corticospinal axons synapse on premotor interneurons in the intermediate zone (Rexed's laminae VII and VIII), and relatively few synapse directly on spinal motoneurons in Rexed's lamina IX. The somata of most corticospinal neurons with monosynaptic connections to motoneurons lie posteriorly in area 4, and their monosynaptic connections are made on the motoneurons of distal limb muscles, whose somata are clustered in the dorsolateral ventral horn. Corticospinal axons from neurons located more anteriorly typically synapse in the ventromedial portion of the ventral horn, where the motoneurons of proximal limb muscles and axial muscles are located. Some corticospinal axons cross the midline spinal gray matter to reach the ventromedial ventral horn ipsilateral to their origin; such doubly decussating corticospinal axons, along with the uncrossed ventral corticospinal tract, may be partly responsible for the relative preservation of trunk and proximal limb movements after unilateral damage to the cortex.

Indirect Pathways to the Spinal Cord Involve Centers in the Brainstem

The corticospinal tract is not the only output pathway through which the cerebral cortex contributes to motor control (Kuypers, 1987). The presence of other, less direct pathways in the macaque has been demonstrated by cutting the medullary pyramid on one side. Stimulation of the cortex on that side still evoked contralateral movements. The somatotopic organization of the cortex in the operated animals was similar to that in normal macaques, although the thresholds for stimulation were raised and distal movements were evoked less often.

Intermixed with corticospinal neurons in layer V of areas 4 and 6 are corticorubral neurons, whose axons project to the red nucleus (RN). Some corticospinal axons also send collaterals to the RN. In addition to cortical inputs, the RN also receives considerable input from the cerebellum (Chapter 31). Many RN neurons, particularly those in the caudal, magno-cellular portion, in turn send their axons across the midline and into the lateral column of the spinal cord, terminating most heavily in the dorsolateral region of the ventral horn. These descending axons from the RN constitute the rubrospinal tract. A pathway thus exists from the cortex to the RN to the spinal cord. The activity of rubrospinal neurons indicates that they play a significant role during voluntary movements of the arm, hand, and fingers. The rubrospinal pathway thus works synergistically with the corticospinal pathway in controlling limb movements.

A second pathway involves neurons scattered in the medial reticular formation of the pons and medulla. The medial reticular formation receives input from cortical motor areas and projects via the ventral column of the spinal cord to the ventral horn, chiefly its ventromedial portion. The axons that make this projection constitute the reticulospinal tract. Although the reticulospinal tract is also part of the medial descending system, the activity of some reticular formation neurons indicates that they also participate in controlling reaching and grasping movements (Baker, 2011).

Organization of the Motor Cortex

The original "motor cortex" is now appreciated to be comprised of several cortical areas. These areas are considered "motor" because (1) they project to other motor

structures, (2) their ablation causes deficits in movement, and (3) their stimulation evokes or alters movements.

The Motor Cortex Is Subdivided into Multiple Cortical Motor Areas

In addition to its descending projections to spinal motoneurons, the motor cortex has cytoarchitectonic features that distinguish it from other regions of the cerebral cortex. The somata of many pyramidal neurons in layer V, the output layer, are exceptionally large (Betz cells). In addition, the neurons of layer IV, the granular layer that receives thalamic input, are very sparse compared to other areas of the neocortex, and hence the motor cortex has been described as dysgranular (area 6) or agranular (area 4) cortex. Based on cytoarchitectonics, myeloarchitectonics, and histochemical features, the motor cortex of macaque monkeys can be subdivided into the cortical motor areas shown in Figure 29.5A. In general, these subdivisions show three mediolaterally oriented strips of cortex, with the primary motor cortex (M1) most posterior and two premotor strips progressively more anterior. Each of these premotor strips has a subdivision on the

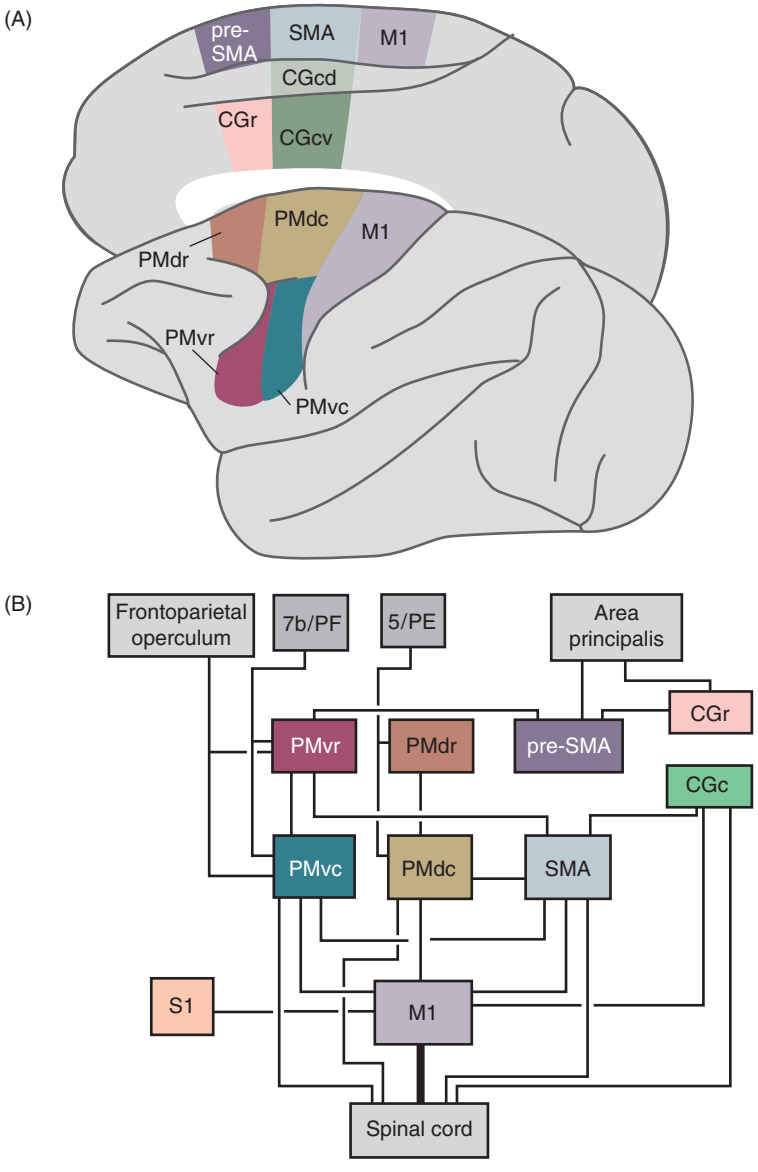


FIGURE 29.5 Cortical motor areas. (A) Diagram of a macaque brain showing current parcellation of cortical motor areas in the frontal lobe. *Modified from Matelli, Luppino, and Rizzolatti (1991).* (B) Connections of the cortical motor areas. Most corticocortical connections are reciprocal. CGc,-cingulate motor area, caudal; CGr,-cingulate motor area, rostral; M1, primary motor cortex; PMdc, premotor cortex, dorsal caudal; PMdr, premotor cortex, dorsal, rostral; PMvc, premotor cortex, ventral, caudal; PMvr, premotor cortex, ventral, rostral; Pre-SMA, presupplementary motor area; SMA supplementary motor area proper. Thin lines to the spinal cord from PMvc, PMdc, SMA, and CGc indicate that corticospinal projections from these areas are not as strong as that from M1.

medial wall of the hemisphere, a second subdivision on the dorsal convexity, and a third on the ventral convexity. In addition to these subdivisions of Brodmann's areas 4 and 6, areas 23 and 24 in the banks of the cingulate sulcus and on the medial surface of the hemisphere in the cingulate gyrus contain at least two additional cortical motor areas. Cortical motor areas differ, not only in their intrinsic composition, but also in their interconnections with other parts of the central nervous system.

Cortical motor areas differ in their connections with the thalamus. M1, PMv, and SMA, for example, connect primarily with VPLo/VLc, area X, and VLo in the thalamus, respectively. These differences are significant because thalamic nuclei VPLo/VLc and area X receive major inputs from the cerebellum, whereas VLo receives major input from the basal ganglia. Thus, information processed by the cerebellum is directed largely to M1 and PMv, whereas information from the basal ganglia is sent largely to SMA.

Cortical motor areas also differ in their connections with one another and with other regions of the cortex (Fig. 29.5B). For example, when horseradish peroxidase (HRP) is injected into the hand region of M1, separate patches of retrogradely labeled neuronal somata are found in PMv, PMd, SMA, CGc, and CGr, indicating that each of these areas sends a separate projection to the hand region of M1. In contrast, other motor cortical regions do not project directly to M1. Instead, PMvr projects to PMvc, PMdr projects to PMdc, and pre-SMA projects to PMvr. The extensive connectivity between motor cortical areas subserves the dynamic visuomotor transformations required to successfully grasp an object (Davare, Kraskov, Rothwell, & Lemon, 2011).

Different cortical motor areas also receive input from different cortical sensory and association areas: M1 receives input from the primary somatosensory area (S1), PMv receives input from visual association area 7b, particularly the anterior intraparietal area (AIP), and PMd receives input from somatosensory association area 5 and the parietal reach region. Because of these inputs, many motor cortex neurons have somatosensory receptive fields. Neurons located caudally in M1 tend to respond to cutaneous modalities, whereas neurons located rostrally in M1 tend to respond to deep modalities. The somatosensory input of an M1 neuron is generally related to the output function of that neuron. For example, caudally located M1 neurons that control the fingers may have cutaneous receptive fields on the fingers (Fig. 29.6A), whereas more rostrally located M1 neurons that control the biceps or triceps may have deep receptive fields in those muscles.

PMv receives both somatosensory and visual input via areas 7b and AIP. Both the somatosensory and the visual receptive fields of PMv neurons tend to be large, and when single PMv neurons receive both types of sensory information, the fields tend to be related (Fig. 29.6B).

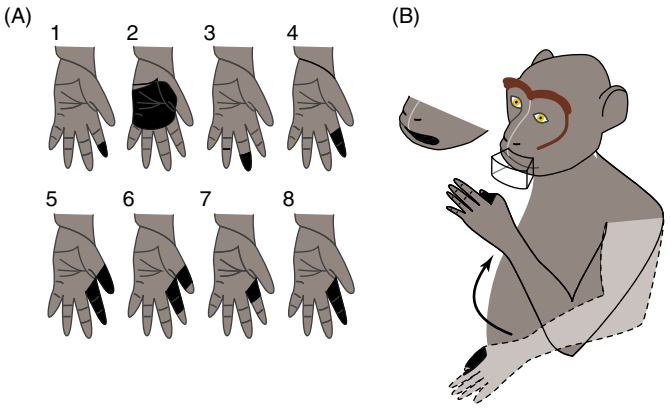


FIGURE 29.6 Sensory receptive fields in M1 and PMv. (A) Black regions show the tactile receptive fields of eight M1 neurons recorded at loci where intracortical microstimulation evoked flexion of the monkey's index and middle fingers. Other neurons at the same loci responded to passive extension of those fingers. *From Rosen and Asanuma (1972).* (B) A single PMv neuron responded to visual stimuli moving near the mouth, to tactile stimulation of the lips and of the skin between the thumb and the index finger, and to flexion of the elbow. *From Rizzolatti, Scandolara, Matelli, and Gentilucci (1981).*

A neuron with a somatosensory receptive field covering the forearm, for instance, may respond to visual stimuli moving near the forearm. Interestingly, if the forearm is moved to a different position, the visual receptive field of the PMv neuron moves with the forearm. The somatosensory inputs to both PMv and M1, plus the visual inputs to PMv, reflect the role of sensory inputs in guiding ongoing movements as they are being generated by cortical motor areas.

Somatotopic Organization in the Motor Cortex Is Not a One-to-One Map of Body Parts, Muscles, or Movements

The organization of the primary motor cortex, M1, has been studied more extensively than that of other cortical motor areas. Within the M1 of primates, the face is represented laterally, the lower extremity (or hindlimb) and tail medially, and the upper extremity (or forelimb) in between (Fig. 29.7). This overall organization of M1 according to major body parts is termed somatotopic. The somatotopic organization of M1 becomes evident in clinical neurology. Lesions on the lateral convexity of human M1 cause weakness or paralysis of the contralateral face, more medial lesions on the convexity affect the contralateral hand and arm, and lesions on the medial wall of the hemisphere affect the leg and foot.

Moreover, distal parts of the extremities and acral parts of the face (lips and tongue) are most heavily represented caudally, whereas proximal parts of the extremities and axial movements are most heavily represented rostrally. Those parts of the body that are used for fine manipulative movements (such as lips, tongue, and fingers in primates)

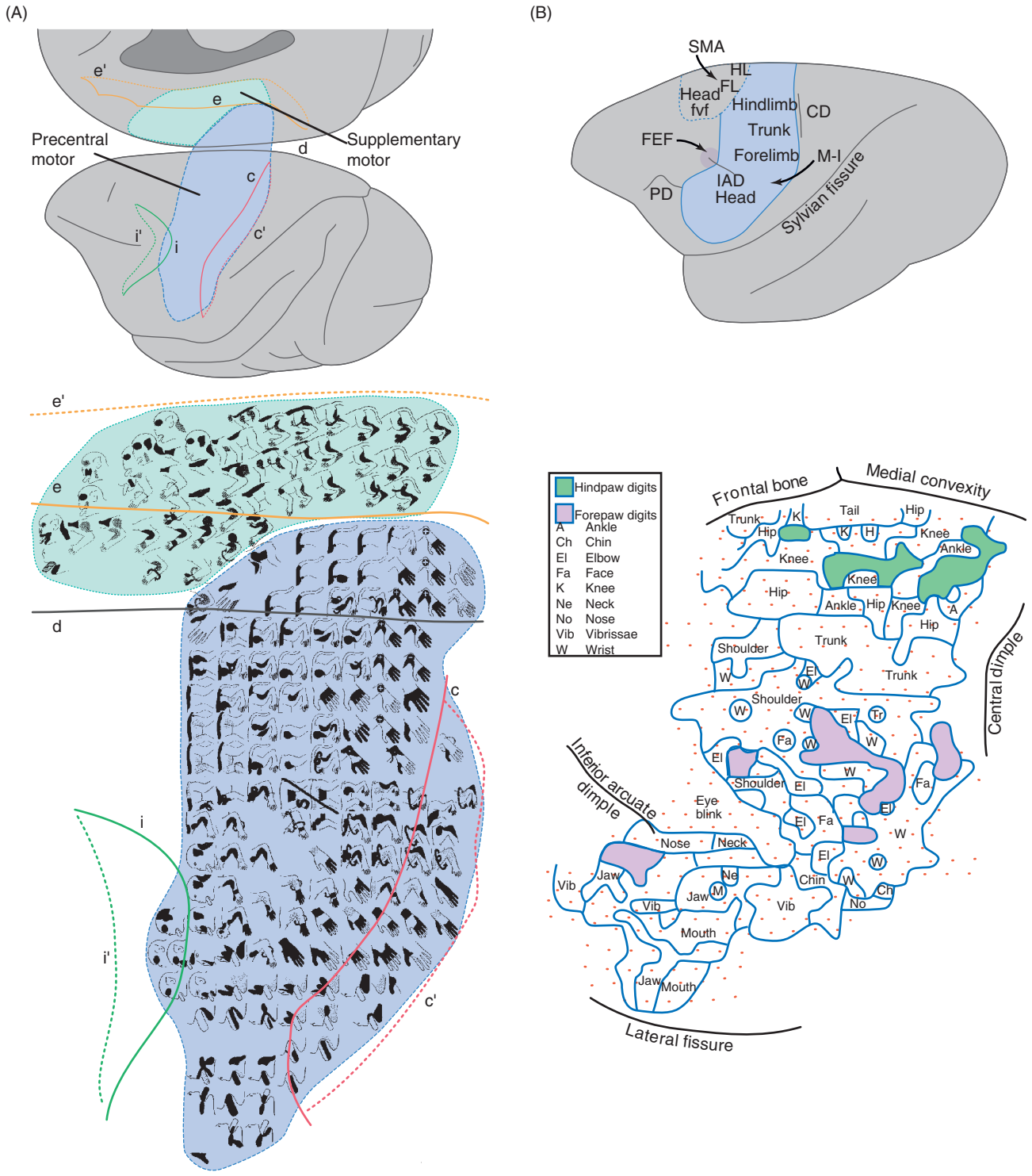


FIGURE 29.7 Somatotopic maps in M1. (A) Map by Woolsey et al. (1952) in which each figurine represents in black and gray the body parts that moved a lot or a little, respectively, when the cortical surface at that site was stimulated. In addition to the primary representation on the convexity, their map shows a secondary representation on the medial surface of the hemisphere, called the supplementary motor area (SMA). As defined in this study, M1 and SMA each included several of the currently defined cortical motor areas. (B) Intracortical microstimulation of M1 in an owl monkey produced this map, consisting of a complex mosaic of different body parts. In this species, the central sulcus is only a shallow dimple, and M1 is entirely on the surface of the hemisphere. Each dot represents a stimulated locus, and lines surround adjacent loci from which movement of the same body part was evoked. Note that the forepaw digits (purple) and the hindpaw digits (green) are represented in multiple areas separated by areas representing nearby body parts. *Modified from Gould, Cusic, Pons, and Kaas (1986).* In both A and B, the inset at the top indicates the region of the frontal lobe enlarged below.

are generally represented over a wider cortical territory than body parts used in gross movements such as ambulation. Penfield, in his cartoon of the motor homunculus (little man), and Woolsey, in his cartoon of the motor simiusculus (little monkey), conveyed this apparent magnification of certain body parts with respect to cortical territory by distorting the size of body parts relative to their normal proportions.

The somatotopic organization of M1 is not, however, a one-to-one mapping of body parts, muscles, or movements. Within the arm representation, for example, the cortical territory representing any particular part of the arm overlaps considerably with the territory representing nearby parts. This overlap results from three features of M1's organization: convergence, divergence, and horizontal interconnection.

M1 outputs arising from a wide cortical territory converge on the spinal motoneuron pool of any given muscle. Convergence of M1 outputs has been shown by delivering intracortical microstimulation while observing the body for evoked movements and/or recording electromyographic (EMG) activity from multiple muscles. Maps of which body parts move or which muscles contract upon threshold stimulation at different cortical sites look like a complex mosaic (Fig. 29.8A). As the stimulation current is increased, the mosaic pieces representing a given muscle coalesce into a larger and larger territory that overlaps more and more with the territories of other muscles. These findings indicate that any given muscle is controlled by a large territory in M1 and that the territories for different muscles overlap (Fig. 29.8B). Using retrograde transneuronal tracing with rabies virus, these large and overlapping cortical territories for different muscles recently have been demonstrated anatomically as well.

A second factor contributing to the overlap of territories in M1 is the divergence of output from single cortical neurons to multiple motoneuron pools. Intracellular staining has shown that a single corticospinal axon may have terminal ramifications within the motoneuron pools of multiple muscles over several segmental levels of the spinal cord (Fig. 29.9A). Moreover, spike-triggered averaging of EMG activity has demonstrated that single M1 neurons can have relatively direct effects on the motoneuron pools of multiple muscles (Fig. 29.9B). These findings indicate that single M1 neurons have output connections that diverge to innervate the spinal motoneuron pools of multiple muscles (Fig. 29.9C).

A third factor contributing to overlapping representation is the horizontal interconnectivity intrinsic to M1. Although most horizontal axon collaterals of neurons in layer V extend only 1–2 mm, some extend across the upper extremity representation, interconnecting even the territories of proximal and distal muscles

Because of convergence, divergence, and horizontal interconnections, neuronal activity is widely distributed

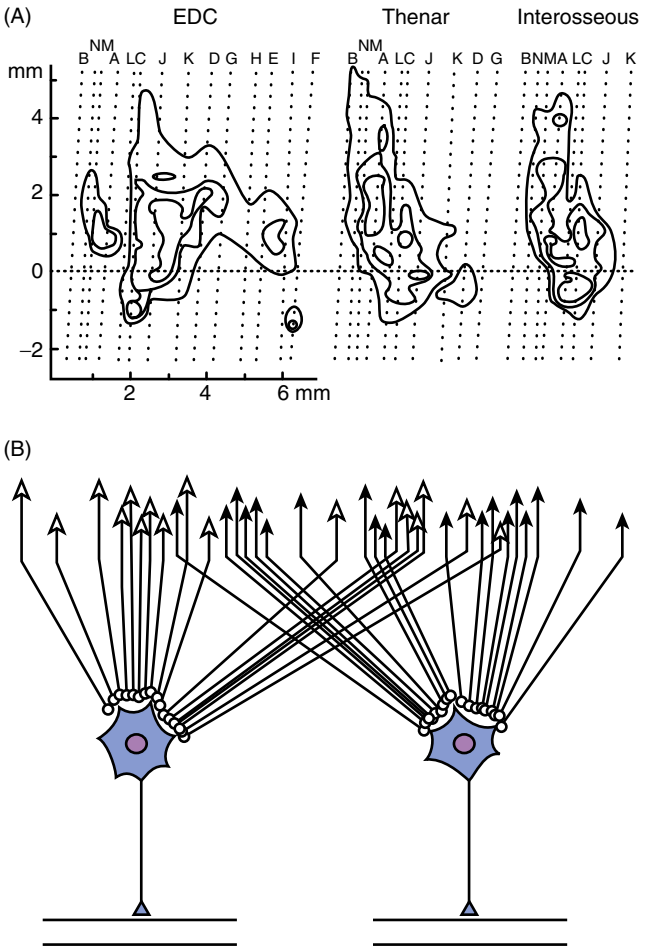


FIGURE 29.8 Convergence of M1 outputs to single muscles. (A) Isothreshold contours show the points at which EMG responses were evoked in three different muscles—extensor digitorum communis (EDC), thenar, and first dorsal interosseus—by intracortical microstimulation in the anterior bank of the central sulcus. (B) Data shown in A indicate that the cortical input to the motoneurons of any muscle originates from a wide territory in M1 and that the cortical territory providing input to a given muscle overlaps extensively with the cortical territory providing input to other muscles in the same part of the body. *Modified from Andersen, Hagan, Phillips, and Powell (1975).*

in M1 during natural movements, even discrete movements of single body parts. In monkeys trained to perform individuated movements of each finger, individual M1 neurons are active during movements of multiple fingers, and neurons throughout the M1 hand area are active during movements of any given finger. Likewise, in humans performing movements of different fingers, activity is distributed over the entire M1 hand representation whether the subject is moving a single finger or the whole hand. Although the entire hand representation may be activated for movement of any given finger, monkey and human studies have shown a tendency for the center of activation during movements of the thumb to be located laterally to that for movements of other fingers, consistent with the

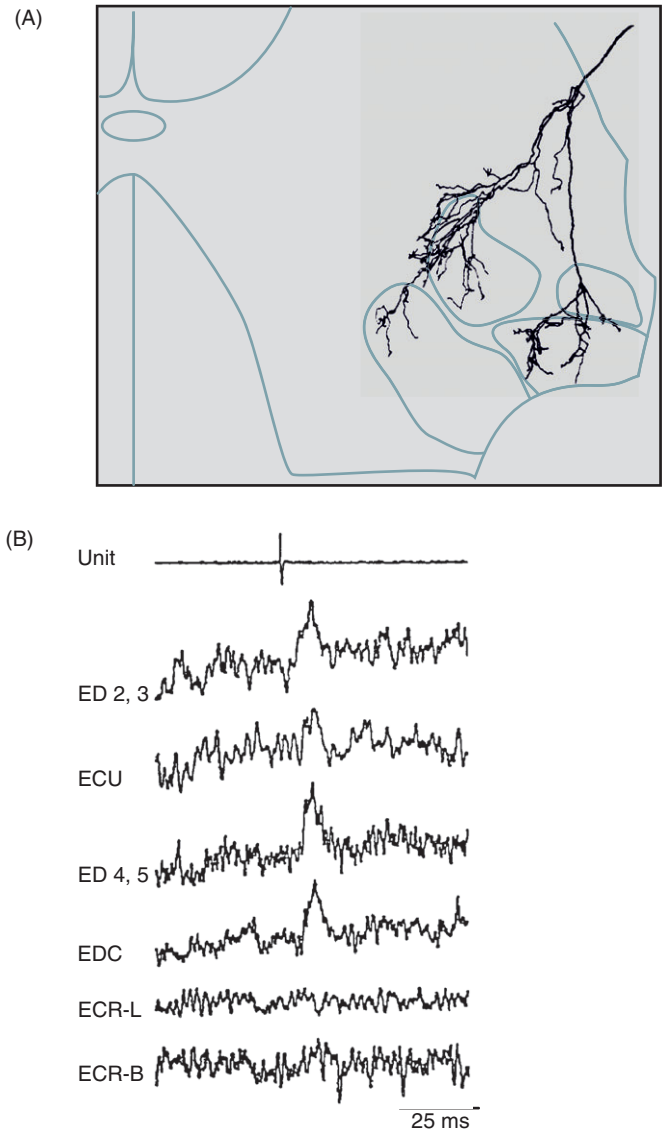


FIGURE 29.9 Divergence of M1 outputs to multiple muscles. (A) Tracing of a single corticospinal axon ramifying in the ventral horn of the spinal cord shows terminal fields in the motoneuron pools of four forearm muscles. *From Shinoda, Yokota, and Futami (1981).* (B) Action potentials in a cortical neuron (top trace) are followed at a fixed latency by peaks of postspike facilitation in EMGs recorded from four of six recorded forearm muscles (lower traces), consistent with monosynaptic excitation of all four motoneuron pools by that cortical neuron. The EMGs are rectified and averaged responses to 7051 action potentials in the cortical neuron. *From Fetz and Cheney (1980).* (C) These anatomic and physiologic findings indicate that the output of single corticospinal neurons often diverges to influence multiple muscles. *From Cheney, Fetz, and Palmer (1985).*

somatotopic orientation of the hand in the classic simiusculus and homunculus.

The somatotopic representation in M1, like that in the primary somatosensory cortex, has a certain degree of plasticity. The cortical territory representing a given muscle can enlarge when nearby body parts are denervated experimentally. Threshold stimulation in the cortical territory that had represented a denervated body part then comes to evoke movements in nearby body parts. Enlargement of a given muscle's cortical territory occurs as well when that muscle is stretched passively or used intensely for a prolonged period. Furthermore, in normal humans who are actively practicing a complex sequence of finger movements, the cortical territory representing a given finger muscle enlarges as the subjects become skilled at performing the sequence. Because such changes can occur within several minutes, they probably are mediated by long-term potentiation and/or depression at existing synapses. This ability of the M1 cortex to reorganize may in part underlie motor recovery seen in humans after damage to M1 or the corticospinal tract.

Control of Voluntary Movements by the Motor Cortex

Several methods have enabled investigators to explore how cortical motor areas contribute to the production and control of voluntary movement in awake, behaving subjects. Whereas microelectrodes record the action potentials of single neurons, measurements made by functional imaging or by recording surface potentials reflect the net activity of thousands of cortical neurons and millions of synapses.

Multiple Cortical Areas Are Active When the Brain Generates a Voluntary Movement

These different methodologies now make it clear that many cortical areas in addition to the primary motor cortex are active during the planning and execution of voluntary movements. During performance of either a simple keypress or a complex movement sequence, for example, functional imaging studies in humans have shown

bilateral activation of the primary sensorimotor hand representation, the supplementary motor area, and the ventrolateral premotor cortex, plus contralateral activation of the dorsolateral premotor cortex and the medial cortex rostral to the SMA. Whereas such techniques provide information on the parts of the cortex that are active in a given situation, studies of electrical potentials provide information on the time course of activation. For simple self-paced movements, cortical electrical potentials over the SMA and M1 begin to change as early as 1 s prior to the movement. As the time of movement onset approaches, the amplitude of these bilateral electrical potential shifts increases (the Bereitschaft potential). At the time of the movement, a further increase in amplitude occurs over the somatotopically appropriate region of M1 contralateral to the moving body part. Given that many cortical areas are active in controlling movement raises the question: What does each area contribute to control?

M1 Neurons Control Movement Kinematics and Dynamics

In one of the earliest studies of single neuron activity in awake behaving animals, Evarts recorded the activity of M1 pyramidal tract neurons in monkeys trained to raise and lower weights using flexion and extension wrist movements (Fig. 29.10). The discharge frequency of many M1 neurons changed systematically in temporal relation to either flexion or extension. A typical flexion-related neuron, for example, began to discharge several hundred milliseconds before wrist flexion began. As the onset of a flexion movement approached, the discharge frequency of the neuron increased, accelerating still more as the flexion movement was made. The neuron was silent during extension. The time course of these movement-related changes in single neuron activity parallels the time course of cortical electrical potential shifts.

Evarts went on to demonstrate that the discharge frequency of M1 pyramidal tract neurons (PTNs) varied in relation to a number of mechanical parameters of the movement the monkey was making. By changing the weights the monkey had to lift, Evarts showed that PTN discharge frequency varied with the force the monkey exerted. Comparison of PTN discharge with simultaneous force recordings revealed that bursts of PTN firing were correlated with sudden increases in the exerted force, indicating a relationship between firing frequency and the rate of change of force.

Subsequent studies revealed that the discharge of M1 neurons can be related to the direction of movement, the position of a particular joint, or the velocity of movement (Kalaska, 2009). Many neurons in the SMA, PMd, and parietal area 5 (which projects to SMA and PMd) also fire in relation to movement direction, and some PMv neurons fire in relation to movement force. Therefore, neurons in several nonprimary cortical motor areas participate with

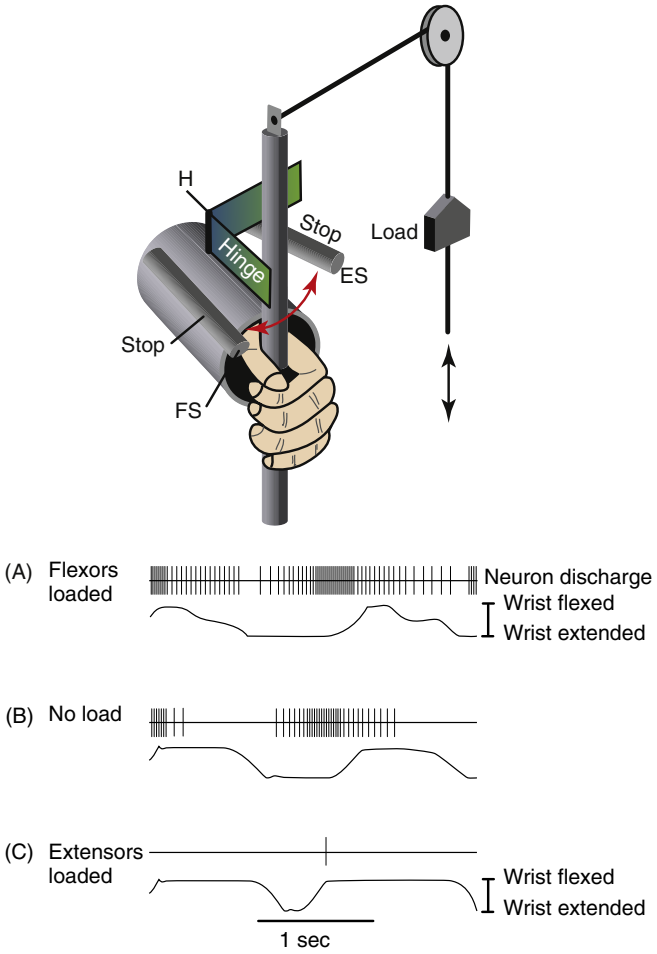


FIGURE 29.10 Discharge of a single M1 neuron in a monkey making flexion and extension wrist movements with wrist flexors loaded (A) and unloaded (B) and with wrist extensors loaded (C). The discharge rate of this neuron was greatest when the monkey used its wrist flexor muscles against a load. *Modified from Evarts (1968).*

those in M1 to control movement parameters such as direction, force, position, and velocity.

Although very strong correlations can be found between the discharge frequency of a given neuron and a particular movement parameter, an M1 neuron does not simply encode a single parameter. Rather, single M1 neurons show various degrees of correlation with direction, force, position, and velocity. Thus, the discharge of a single neuron in the motor cortex may influence several movement parameters.

Conversely, any given parameter or other feature of a movement is represented not by the discharge of a single M1 neuron, but by the ensemble activity of a large population of cortical neurons. Although a single M1 neuron typically discharges during reaching movements in several directions, for example, when the activity of many M1 neurons is combined appropriately, the population of M1 neurons can be shown to represent movement direction

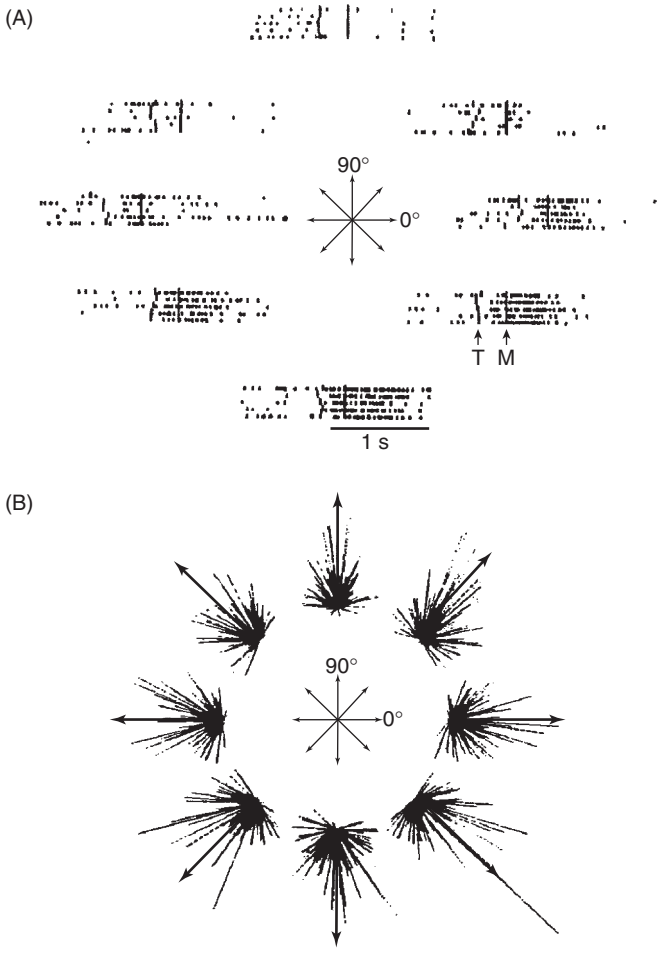


FIGURE 29.11 (A) Discharge of a single M1 neuron before and during arm movements in a monkey. Movements (represented by arrows) started from the same central point and ended at eight different points on a circle. The eight rasters show that the activity of this neuron was related to movements in four of the eight directions. The neuron discharged most intensely for movements down and to the right and was inhibited during movements up and to the left. (B) For each of the eight movements, the discharge of each M1 neuron is shown as a line pointing in the preferred direction of the neuron. Each line starts at the movement end point, and its length is proportional to the intensity of the discharge of that neuron during movement in that direction. Although the discharge of single neurons rarely identified any single movement direction with accuracy, the population vectors (arrows) summing the discharge of an ensemble of M1 neurons adequately specify each of the eight movement directions. *From Georgopoulos (1988).*

precisely (Fig. 29.11). Researchers have correlated the force, rate of change of force, position, and velocity of movements more accurately with the summed, weighted activity of a number of simultaneously recorded M1 neurons than with the discharge of any single neuron (see Box 29.4).

Cortical Motor Areas Prepare Voluntary Movements Based on a Variety of Cues

Cortical areas outside the primary motor cortex seem to be especially concerned with using a wide variety of sensory and other information as "cues" to select and guide movements. Insight into these cortical processes has been obtained by separating in time a cue instructing which movement to make from a second cue to execute the instructed movement. This creates a period between the two cues during which the subject knows what movement to make, but is not making the movement per se. During such an instructed delay period, many neurons in M1, SMA, and PMd discharge at their highest rate while the subject waits to move in a particular direction (Fig. 29.12). Such directional delay period activity is more common in PMd and SMA than in M1, where activity during movement execution predominates. During the delay between instruction and trigger, PMd, SMA, and other areas appear to store information on the direction of the impending movement. Indeed, neurons sometimes discharge in error during the delay, as if the monkey has seen a cue other than the one actually given and is preparing to move in the wrong direction. When such error discharges occur, the monkey often does make the wrong movement. The delay period discharge of such neurons thus represents stored information, not about which cue the monkey has seen, but rather about what movement the monkey will make.

The direction instructed usually is the same as the direction of the actual movement. To pick up a pencil, for example, you look at the pencil and then move your hand to the same place. Your brain transforms the visual location of the pencil into the location to which your hand is moved. Insight into how cue direction is transformed into movement direction has come from tasks in which these two features were experimentally dissociated, as if you looked at a pencil in a mirror and then reached to pick up the real pencil instead of the mirror image. To dissociate cue direction and movement direction experimentally, researchers have trained monkeys to perform mental rotation tasks in which a cue at one location instructs a movement to a different location.

Such tasks have been used to study neurons in the area principalis of the dorsolateral frontal lobe, in PMd, in SMA, and in M1. In each of these cortical areas, the activity of some neurons correlates best with the direction of the instructional cue, whereas the activity of other neurons correlates best with the direction of the actual movement. In general, from the area principalis, through the PMd and SMA to M1, the percentage of cue-related neurons decreases and the percentage of movement-related neurons increases. Moreover, as time progresses from the appearance of the instructional cue to the execution of the movement, the discharge of the neuronal populations encode cue direction initially and movement direction subsequently. The transformation of cue direction into movement direction thus involves many cortical motor areas and progresses throughout the reaction time of the subject.

BOX 29.4

MOTOR NEUROPROSTHETICS

The fact that a subject's movements can be decoded from populations of neurons, combined with the recently developed ability to implant devices in the brain that record populations of neurons simultaneously, has led to the development of brain-computer interfaces (BCIs) and brain machine interfaces (BMIs). These are devices in which a number of simultaneously recorded neurophysiological signals are decoded as a population, and the decoded output then is used to control either a cursor on a computer screen (BCI) or a physical device such as a robotic arm (BMI). While recordings of neuron spikes generally provide the best decoding, other types of neurophysiological signals—local field potentials recorded from penetrating microelectrodes (LFPs), recordings made from various sites on the surface of the brain (electrocorticographic, ECoG), or recordings obtained from the scalp (electroencephalographic, EEG)—all can provide decodable information about a subject's movements. With current systems, subjects can control the movement of an artificial arm in three-dimensional space, and can open and close the hand. The hope is that with further development of such neuroprosthetic systems, amputees may be able to have dexterous control of a prosthetic arm and hand, and paralyzed patients may be able to work on a computer or move in a normal fashion by driving electrical stimulation of their own, otherwise intact muscles (Lebedev & Nicolelis, 2006).

In addition to the potential for useful devices controlled directly from the brain, the field of neuroprosthetics is expanding our understanding of the motor cortex. In a typical BCI session, neurophysiological activity is recorded while a normal subject performs arm movements that control a cursor on a computer screen, like when you move a computer mouse to control a cursor. A decoding algorithm then is derived from the recorded data to predict which patterns of neurophysiological activity correspond to which movements. Next, the decoding algorithm is used to drive a computer cursor. If in this "brain control" mode the subject again is asked to move the cursor to different locations on the screen, however, a number of interesting things often happen. The tuning relationship between neurophysiological activity and movement may change, and the subject may stop moving entirely, even though the computer cursor continues to move under direct control by the cortical activity. Exactly how the cortical activity becomes dissociated from movement of the limb (Schieber, 2011), and how the cortical activity adapts to control the BCI, are currently questions under investigation (Green & Kalaska, 2011).

Marc H. Schieber

References

Green, A. M., & Kalaska, J. F. (2011). Learning to move machines with the mind. *Trends in Neurosciences, 34,* 61–75.

Lebedev, M. A., & Nicolelis, M. A. (2006). Brain-machine interfaces: Past, present and future. *Trends in Neurosciences, 29,* 536–546.

Schieber, M. H. (2011). Dissociating motor cortex from the motor. *Journal of Physiology, 589,* 5613–5624.

Some Cortical Motor Areas are Involved in Movements in Response to Either Internal or External Cues

Another aspect of movement preparation that differentially involves certain cortical motor areas has to do with whether the instructions about what to do come from internally remembered or externally delivered cues. In one experiment, for example, monkeys were trained to touch three of four pads in randomly selected sequences that were cued when the pads were lit in sequence. Once a monkey was accustomed to a given sequence, the lights were gradually dimmed until the monkey was performing the correct sequence based only on internally remembered cues. After several of these remembered trials, the pads were lit in a different sequence for several trials. Whereas neurons in M1 showed similar discharge rates during the internally remembered and externally cued trials for a given sequence, SMA neurons were generally more active during the internally remembered trials, and PMv neurons were generally more active during the externally cued trials (Fig. 29.13).

Summary

Axons from the cerebral cortex and from the red nucleus in the brainstem descend in pathways located dorsolaterally in the spinal cord to control voluntary movements. In addition to corticospinal and rubrospinal pathways, the motor cortex also sends projections to the red nucleus and to the pontomedullary reticular formation, providing additional pathways for voluntary control. Multiple cortical motor areas—distinguished by their inputs, their cytoarchitecture, and their interconnections with other cortical areas and with the thalamus—make different contributions to the control of voluntary movements. The most detailed motor map is found in the

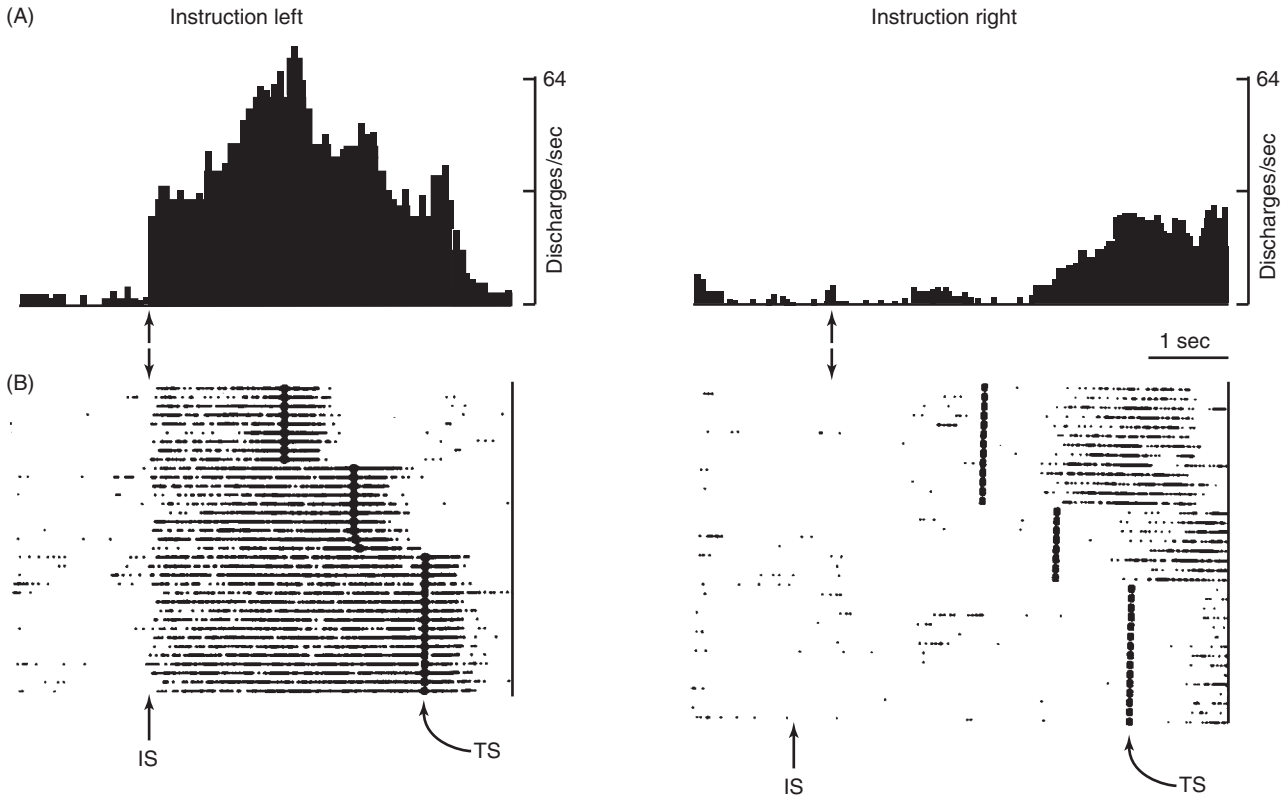


FIGURE 29.12 Directional delay period activity in a PM neuron. (A) As a monkey performed a delayed-reaction paradigm, this neuron began to discharge shortly after receiving instructions (IS) to perform a leftward movement. Discharge continued until after the monkey had subsequently received a separate triggering signal (TS, which occurred at three different time intervals after the IS) and performed the movement. During the delay between IS and TS, while the monkey did not move, the discharge of the neuron encoded the direction of the instruction, the direction of the impending movement, or both. (B) When the instruction was for rightward movement, this neuron did not discharge until after the movement had been made, presumably as the monkey was then preparing to move back to its original position. *From Wise and Strick (1996).*

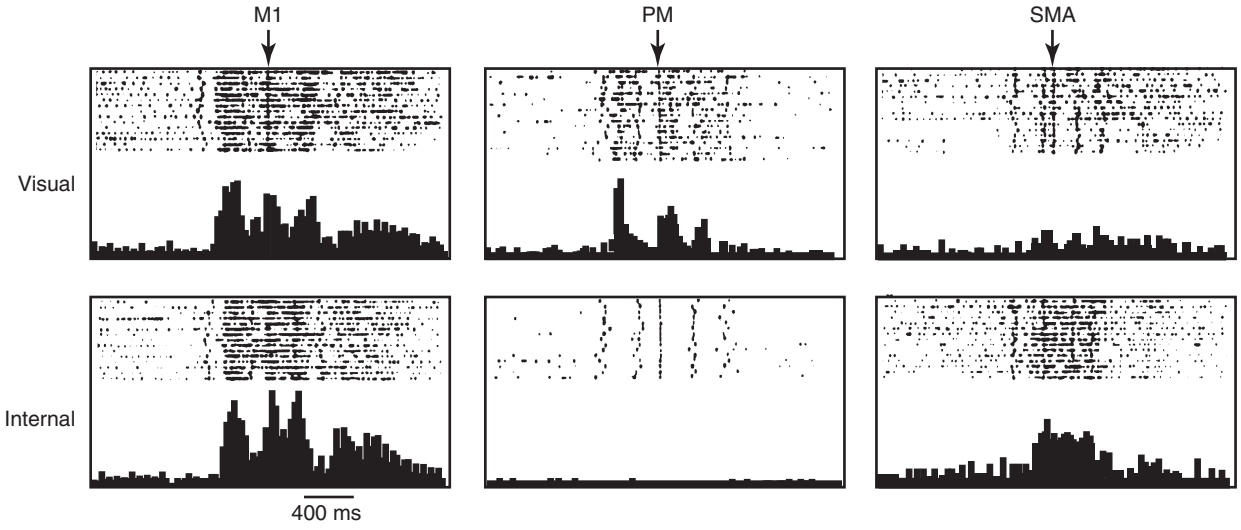


FIGURE 29.13 Activity of three neurons—one in M1, one in PM, and one in SMA—recorded as a monkey pressed three buttons in sequence. The sequence was first cued visually by lighting the buttons and was then cued internally. The M1 neuron showed similar activity whether the monkey performed from visual or internal cues. The PM neuron, however, was much more active in response to visual than internal cues, whereas the opposite was true for the SMA neuron. *Modified from Mushiake, Inase, and Tanji (1991).*

primary motor cortex, M1, but the somatotopic organization is limited by convergence and divergence in the corticospinal projection. Populations of M1 neurons are also in most direct control of movement kinematics and dynamics. Other cortical motor areas participate differentially in the selection of, and preparation for, voluntary movements based on a variety of internal and external cues.

SUMMARY

The motor cortex, the red nucleus, the pontomedullary reticular formation, and the vestibular nuclei each send major axonal projections descending from the brain to the spinal cord to control bodily movements. Although these pathways normally provide seamless movement control, different contributions of the medial and lateral descending pathways have been identified experimentally.

The medial system—vestibulospinal and reticulospinal—receives sensory input from the vestibular apparatus and mediates postural responses that keep the head and body stabilized for stance and gait. Additional visual and proprioceptive inputs that reach the vestibular and reticular nuclei via brainstem pathways or after processing in the cerebellum (Chapter 31) also influence postural control. Context-dependent strategies established via centers including the cerebellum and motor cortex further adapt postural responses to the needs of particular complex situations.

The lateral system—corticospinal and rubrospinal—receives, in addition to somatosensory inputs, information processed by the cerebellum (Chapter 31), the basal ganglia (Chapter 30), and association cortical areas to mediate voluntary movements of the face and extremities. Different cortical motor areas receive different portions of these inputs and therefore make different contributions to controlling voluntary movements. The most elaborate somatotopic representation is found in the primary motor cortex, M1. While the convergence, divergence and horizontal interconnections of the intrinsic organization of M1 result in considerably distributed activation during voluntary movements, the same substrates enable substantial plasticity. Whereas M1 neurons most directly control the kinematics and dynamics of movement execution, other cortical motor areas participate in selecting which movement(s) to make on the basis of external cues and internal states.

References

Andersen, P., Hagan, P. J., Phillips, C. G., & Powell, T. P. (1975). Mapping by microstimulation of overlapping projections from area 4 to motor units of the baboon's hand. *Proceedings of the Royal Society of London B Biological Sciences, 188,* 31–36.

Angelaki, D. E., & Cullen, K. E. (2008). Vestibular system: The many facets of a multimodal system. *Annual Review of Neuroscience, 31,* 125–150.

Bagnall, M. W., McElvain, L. E., Faulstich, M., & du Lac, S. (2008). Frequency-independent synaptic transmission supports a linear vestibular behavior. *Neuron, 60,* 343–352.

Baker, S. N. (2011). The primate reticulospinal tract, hand function and functional recovery. *Journal of Physiology, 589,* 5603–5612.

Baker, J., Goldberg, J., & Peterson, B. (1985). Spatial and temporal response properties of the vestibulocollic reflex in decerebrate cats. *Journal of Neurophysiology, 54,* 735–756.

Baloh, R., & Honrubia, V. (1990). *Clinical neurophysiology of the vestibular system* (2nd ed.). Philadelphia, PA: F. A. Davis.

Cheney, P. D., Fetz, E. E., & Palmer, S. S. (1985). Patterns of facilitation and suppression of antagonist forelimb muscles from motor cortex sites in the awake monkey. *Journal of Neurophysiology, 53,* 805–820.

Davare, M., Kraskov, A., Rothwell, J. C., & Lemon, R. (2011). Interactions between areas of the cortical grasping network. *Current Opinion in Neurobiology, 21,* 565–570.

Evarts, E. V. (1968). Relation of pyramidal tract activity to force exerted during voluntary movement. *Journal of Neurophysiology, 31,* 14–27.

Fetz, E. E., & Cheney, P. D. (1980). Postspike facilitation of fore-limb muscle activity by primate corticomotoneuronal cells. *Journal of Neurophysiology, 44,* 751–772.

Georgopoulos, A. P. (1988). Neural integration of movement: Role of motor cortex in reaching. *FASEB Journal, 2,* 2849–2857.

Gould, H. J., Cusick, C. G., Pons, T. P., & Kaas, J. H. (1986). The relationship of corpus callosum connections to electrical stimulation maps of motor, supplementary motor, and the frontal eye fields in owl monkeys. *The Journal of Comparative Neurology, 247,* 297–325.

Horak, F., & Nashner, L. (1986). Central programming of postural movements: Adaptation to altered support-surface configurations. *Journal of Neurophysiology, 55,* 1369–1381.

Horak, F., & Shupert, C. (1994). Role of the vestibular system in postural control. In S. Herdman (Ed.), *Vestibular rehabilitation*. Philadelphia, PA: F. A. Davis.

Kalaska, J. F. (2009). From intention to action: Motor cortex and the control of reaching movements. *Advances in Experimental Medicine and Biology, 629,* 139–178.

Kolkman, K. E., McElvain, L. E., & du Lac, S. (2011). Diverse precerebellar neurons share similar intrinsic excitability. *Journal of Neuroscience, 31,* 16665–16674.

Kuypers, H. G. (1987). Some aspects of the organization of the output of the motor cortex. *Ciba Foundation Symposium, 132,* 63–82.

Lawrence, D. G., & Kuypers, H. G. (1968a). The functional organization of the motor system in the monkey. I. The effects of bilateral pyramidal lesions. *Brain, 91,* 1–14.

Lawrence, D. G., & Kuypers, H. G. (1968b). The functional organization of the motor system in the monkey. II. The effects of lesions of the descending brain-stem pathways. *Brain, 91,* 15–36.

Matelli, M., Luppino, G., & Rizzolatti, G. (1991). Architecture of superior and mesial area 6 and the adjacent cingulate cortex in the macaque monkey. *Journal of Comparative Neurology, 311,* 445–462.

Mushiake, H., Inase, M., & Tanji, J. (1991). Neuronal activity in the primate premotor, supplementary, and precentral motor cortex during visually guided and internally determined sequential movements. *Journal of Neurophysiology, 66,* 705–718.

Nashner, L. (1982). Adaptation of human movement to altered environments. *Trends in Neurosciences, 5,* 358–361.

Peterson, B., Goldberg, J., Bilotto, G., & Fuller, J. (1985). Cervico-collic reflex: Its dynamic properties and interaction with vestibular reflexes. *Journal of Neurophysiology, 54,* 90–109.

Rizzolatti, G., Scandolara, C., Matelli, M., & Gentilucci, M. (1981). Afferent properties of periarcuate neurons in macaque monkeys. II. Visual responses. *Behavioural Brain Research, 2,* 147–163.

Roberts, T. (1967). *Neurophysiology of postural mechanisms*. New York, NY: Plenum Press.

Rosen, I., & Asanuma, H. (1972). Peripheral afferent inputs to the forelimb area of the monkey motor cortex: Input-output relations. *Experimental Brain Research, 14,* 257–273.

Schor, R., Kearney, R., & Dieringer, N. (1988). Reflex stabilization of the head. In B. Peterson & F. Richmond (Eds.), *Control of head movement*. New York, NY: Oxford University Press.

Shinoda, Y., Yokota, J., & Futami, T. (1981). Divergent projection of individual corticospinal axons to motoneurons of multiple muscles in the monkey. *Neuroscience Letters, 23,* 7–12.

Wilson, V. & Melvill-Jones, G. (1979). *Mammalian vestibular physiology*. New York, NY: Plenum Press.

Wise, S. P., & Strick, P. L. (1996). Anatomical and physiological organization of the non-primary motor cortex. *TINS, 7,* 442–446.

Woolsey, C. N., Settlage, P. H., Meyer, D. R., Sencer, W., Hamuy, T. P., & Travis, A. M. (1952). Patterns of localization in precentral and "supplementary" motor areas and their relation to the concept of a premotor area. *Research Publications - Association for Research in Nervous and Mental Disease, 30,* 238–264.

Marc H. Schieber and James F. Baker

This page intentionally left blank

CHAPTER

30

The Basal Ganglia

The basal ganglia and the cerebellum are large subcortical components of the motor system. They act in parallel, influencing both cortical and brainstem motor areas to ensure efficient, focused, and well-coordinated movements. Although they communicate with the same areas of cerebral cortex and with each other, they have distinct functions and mechanisms of action. They communicate with motor areas of cerebral cortex privately using separate thalamic nuclei. The outputs have opposite signs: the basal ganglia output is inhibitory and the cerebellar output is excitatory. The cerebellum receives direct input from spinal cord ascending pathways and from the vestibular system, whereas the basal ganglia receives the bulk of its sensory information from cerebral cortex. In human diseases, basal ganglia dysfunction results in changes in the slow voluntary movements or the development of involuntary movements; cerebellar dysfunction results in impaired coordination. In addition to their motor functions, both basal ganglia and cerebellum play roles in nonmotor behavior, including cognition, emotion, and possibly others. The basal ganglia are discussed in this chapter and the cerebellum in the next (Chapter 31).

The basal ganglia comprise several interconnected nuclei in the forebrain, diencephalon, and midbrain (Fig. 30.1). They are an important part of the motor system, and motor control is a large part of basal ganglia function: (1) the largest portion of basal ganglia inputs and outputs are connected with motor areas; (2) the discharge of many basal ganglia neurons correlates with movement; and (3) basal ganglia lesions cause severe movement abnormalities. In view of these features, this chapter concentrates on the motor functions of the basal ganglia. Participation of the basal ganglia in nonmotor functions is discussed after the basic framework of basal ganglia function is presented for motor control. The anatomy of the basal ganglia is discussed first, since it is anatomy that provides the infrastructure for the physiology. Next we will consider the activity of the individual components of the basal ganglia during movement. Then we will review the effect of placing a lesion in one basal ganglia component while leaving the rest of the nervous system intact. The combined strategy of measuring the activity of a structure during movement and then determining the effect of a selective lesion on that movement has been fruitful in studying the function of the motor system and understanding the effects of human basal ganglia diseases. After presentation of current models of basal ganglia motor function, the chapter concludes with a discussion of basal ganglia participation in cognitive function.

ANATOMY OF THE BASAL GANGLIA

The basal ganglia receive a broad spectrum of cortical inputs. The information conveyed to the basal ganglia is processed to produce a focused output to areas of the frontal lobes and brainstem that are involved in the planning and production of movement. The basal ganglia output is inhibitory, which means that an *increase* in basal ganglia output leads to a *reduction* in the activity of its targets. The fact that the basal ganglia output is inhibitory to other motor mechanisms is key to understanding its normal function.

At first glance, the anatomy of the basal ganglia may seem confusing. There are several component nuclei at different levels in the brain, and two of the nuclei [substantia nigra (SN) and globus pallidus] are divided into functionally different components. The names of some structures are different in primates than in other mammals. Furthermore, the terminology has changed over the years so that

© 2013 Elsevier Inc. All rights reserved.

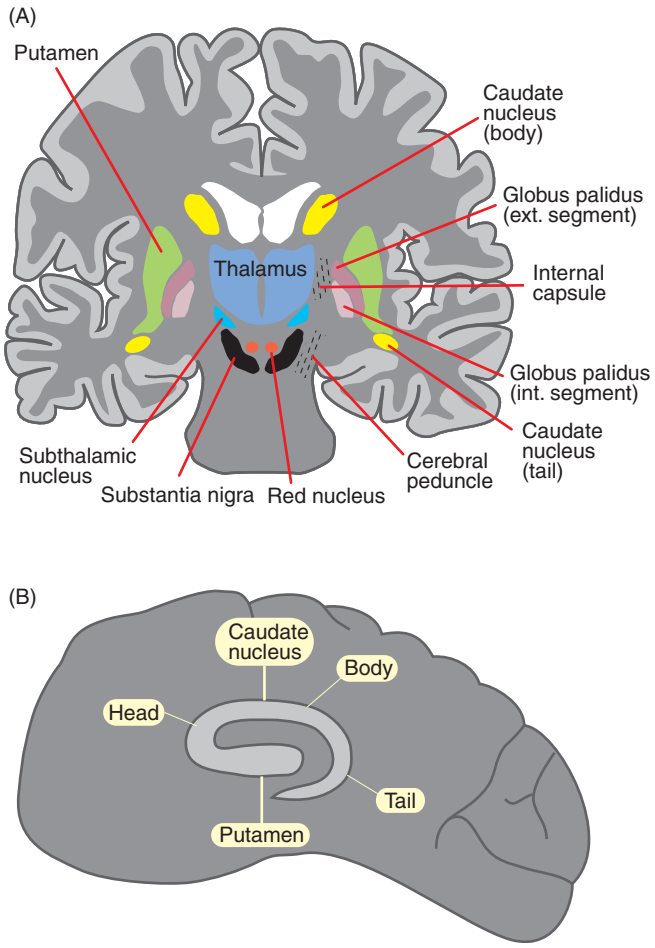


FIGURE 30.1 Location of basal ganglia in the human brain. (A) Coronal section. (B) Parasagittal section.

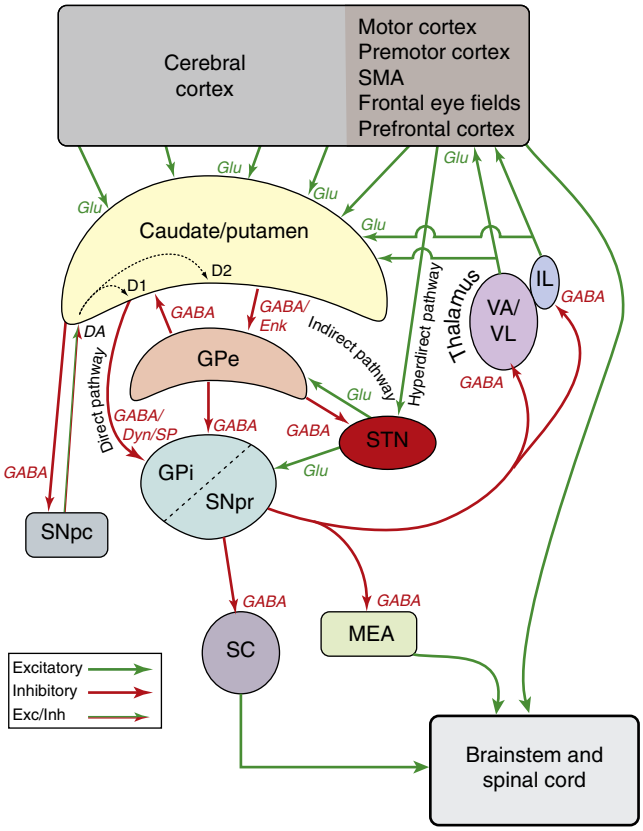


FIGURE 30.2 Simplified schematic diagram of basal ganglia circuitry. Excitatory connections are indicated by green arrows, inhibitory connections by red arrows, and the modulatory dopamine projection is indicated by a red and green arrow. GPe, globus pallidus, external segment; GPi, globus pallidus, internal segment; IL, intralaminar thalamic nuclei; MD, mediodorsal thalamic nucleus; MEA, midbrain extrapyramidal area; SC, superior colliculus; SNpc, substantia nigra pars compacta; SNpr, substantia nigra, pars reticulata; STN, subthalamic nucleus; VA, ventral anterior thalamic nucleus; VL, ventral lateral thalamic nucleus; DA, dopamine (with D1 and D2 receptor subtypes); Dyn, dynorphin; Enk, enkephalin; GABA, γ-aminobutyric acid; Glu, glutamate; SP, substance P.

historically an individual structure has been known by more than one name. However, with consistent use of modern terminology and a functional context in which to place the anatomy, the organization of basal ganglia can be learned readily. Once the organization is learned, many aspects of normal function and the response to injury become easier to understand.

The basal ganglia include the striatum (caudate, putamen, nucleus accumbens); the subthalamic nucleus (STN); the globus pallidus (internal segment—GPi; external segment—GPe; and ventral pallidum—VP); and the substantia nigra (pars compacta—SNpc; and pars reticulata—SNpr) (Fig. 30.2). The striatum and STN receive the majority of their inputs from outside the basal ganglia. Most of those inputs come from the cerebral cortex, but thalamic nuclei also provide strong inputs to striatum. There are no direct inputs from peripheral sensory or motor systems. The bulk of the outputs from basal ganglia arise from GPi and SNpr and are inhibitory to thalamic nuclei and to the brainstem. There are no direct outputs from basal ganglia to spinal motor circuitry.

The Striatum Receives Most of the Inputs to the Basal Ganglia

The striatum is located in the forebrain and comprises the caudate nucleus and putamen (neostriatum) and nucleus accumbens (ventral striatum). This section focuses on the neostriatum (see Box 30.3 and Chapter 41 for a discussion of the ventral system of the basal ganglia). Embryologically, the striatum develops from a basal region of the lateral telencephalic vesicle. It is named the striatum because axon fibers passing through the striatum give it a striped appearance. In rodents, the caudate and putamen form a single structure (caudatoputamen) with fibers of the internal capsule coursing through, but in carnivores and primates, the caudate and putamen are separated by the internal capsule (Fig. 30.1A).

Four major types of neurons have been described in the striatum, based on the size of the cell body, presence or absence of dendritic spines in individual neurons that have been stained with the Golgi technique or filled with horseradish peroxidase, and other staining properties. The first and by far the most numerous neuron type is the *medium spiny neuron* (Fig. 30.3). Medium spiny neurons make up 90 to 95% of the total number of striatal neurons. They utilize γ-aminobutyric acid (GABA) as a transmitter and constitute the output of the striatum, sending axonal projections to the GP and SN. Medium spiny neurons have large dendritic trees that span 200 to 500 μm. They also have extensive local axon collaterals that inhibit neighboring striatal neurons.

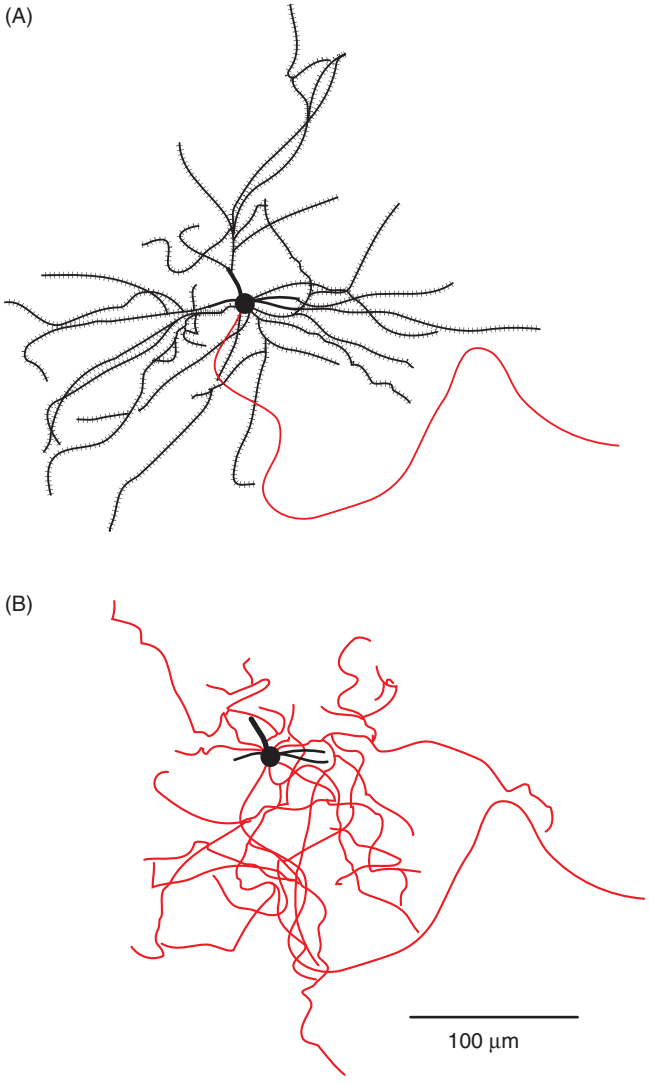


FIGURE 30.3 Two representations of a striatal medium spiny neuron that has been filled with HRP. (A) The soma and dendritic tree with numerous dendritic spines. The thin process is the axon, which has been drawn without its collaterals. (B) The same neuron drawn to show the axonal collaterals, which branch extensively within the same field as the dendritic tree. *From Wilson and Groves (1980).*

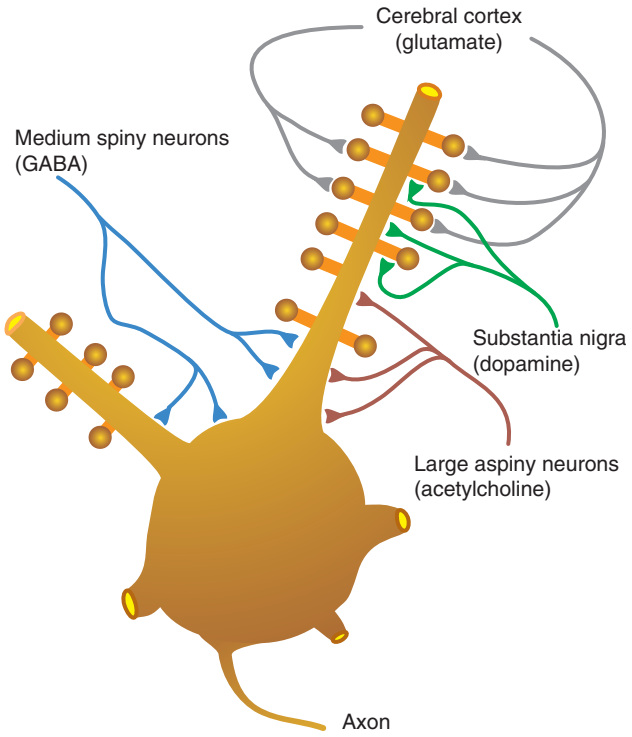


FIGURE 30.4 Pattern of termination of afferents on a medium spiny neuron. The soma and the proximal dendrites with their spines are shown.

The medium spiny neurons are morphologically homogeneous, but chemically heterogeneous, as discussed later. The second, *large aspiny neurons*, make up 1 to 2% of the striatal population. They are interneurons and use acetylcholine (ACh) as a neurotransmitter. They have extensive axon collaterals within the striatum that terminate on medium spiny neurons. The third type, *small aspiny neurons*, make up 2 to 5% of the striatal population. They stain positive for parvalbumin and use GABA as a neurotransmitter. The fourth type is the *medium aspiny neuron*. These are also interneurons and are thought to use somatostatin as a neurotransmitter.

The striatum receives excitatory input from nearly all of the cerebral cortex. The cortical input uses glutamate as its neurotransmitter and terminates largely on the heads of the dendritic spines of medium spiny neurons (Fig. 30.4). The projection from the cerebral cortex to the striatum has a roughly topographical organization. For example, the somatosensory and motor cortex project to the posterior putamen and the prefrontal cortex projects to the anterior caudate. Within the somatosensory and motor projection to the striatum, there is a preservation of somatotopy. It has been suggested that the topographic relationship between the cerebral cortex and the striatum provides a basis for the segregation of functionally different circuits in the basal ganglia (Kelly & Strick, 2004) (Fig. 30.5). These circuits include

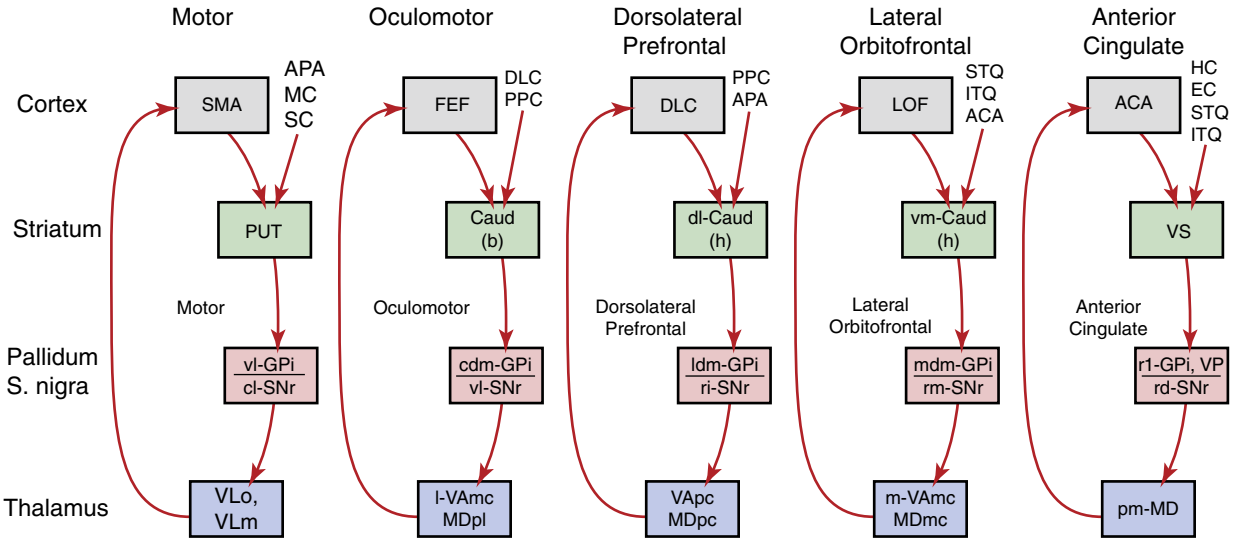


FIGURE 30.5 Hypothetical parallel segregated circuits connecting the basal ganglia, thalamus, and cerebral cortex. The five circuits are named according to the primary cortical target of the output from the basal ganglia: motor, oculomotor, dorsolateral prefrontal, lateral orbitofrontal, and anterior cingulate. ACA, anterior cingulate area; APA, arcuate premotor area; CAUD, caudate; b, body; h, head; DLC, dorsolateral prefrontal cortex; EC, entorhinal cortex; FEF, frontal eye fields; GPi, internal segment of globus pallidus; HC, hippocampal cortex; ITG, inferior temporal gyrus; LOF, lateral orbitofrontal cortex; MC, motor cortex; MDpl, medialis dorsalis pars paralarnellaris; MDme, medialis dorsalis pars magnocellularis; MDpc, medialis dorsalis pars parvocellularis; PPC, posterior parietal cortex; PUT, putamen; SC, somatosensory cortex; SMA, supplementary motor area; SNr, substantia nigra pars reticulate; STG, superior temporal gyrus; VAmc, ventralis anterior pars magnocellularis; Vapc, ventralis anterior pars parvocellularis; VLm, ventralis lateralis pars medialis; VLo, ventralis lateralis pars oralis; VP, ventral pallidum; VS, ventral striatum, cl, caudolateral; cdm, caudal dorsomedial; d1, dorsolateral; 1, lateral; 1dm, lateral dorsomedial; m, medial; mdm, medial dorsomedial; pm, posteromedial; rd, rostrodorsal; rl, rostrolateral; rm, rostromedial; vm, ventromedial; vl, ventrolateral.

somatomotor, oculomotor, cognitive, and limbic connections. Within each circuit there appear to be subcircuits such that the primary motor cortex and premotor cortex have nonidentical connections with basal ganglia structures. Likewise, dorsolateral and orbitofrontal circuits have distinct connectivity patterns.

Although the topography and somatotopy imply a certain degree of parallel organization, there are also convergence and divergence in the corticostriatal projection. The large dendritic fields of medium spiny neurons allow them to receive input from adjacent projections, which arise from different areas of cortex. Inputs from more than one cortical area overlap, and input from a single cortical area projects divergently to multiple striatal zones (Fig. 30.6). This convergent and divergent organization provides an anatomical framework for the integration and transformation of information from several areas of the cerebral cortex.

In addition to cortical input, medium spiny striatal neurons receive a number of other inputs, including (1) excitatory glutamatergic inputs from intralaminar and ventrolateral nuclei of the thalamus; (2) cholinergic input from large aspiny neurons; (3) GABA, substance P, and enkephalin input from adjacent medium spiny striatal neurons; (4) GABA input from small interneurons; and (5) a large input from dopamine (DA)-containing neurons in the SNpc. The DA input is of particular interest because of its role in Parkinson disease (see Box 30.1).

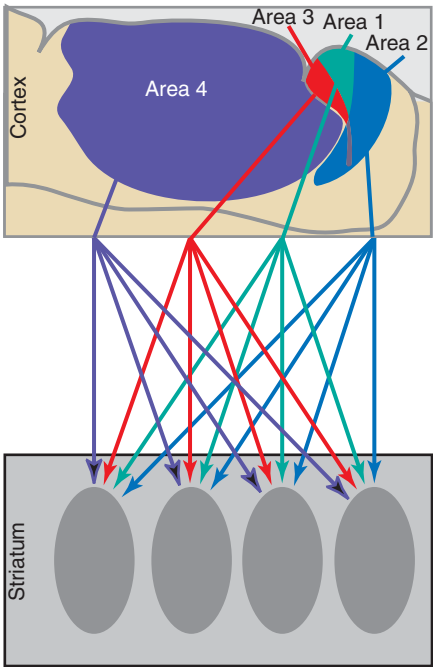


FIGURE 30.6 Schematic representation of the sensorimotor cortical projection to the striatum from arm areas in the somatosensory cortex (areas 1, 2, and 3) and motor cortex (area 4). Note that each cortical area projects to several striatal zones and several functionally related cortical areas project to a single striatal zone. *After Flaherty and Graybiel (1991).*

BOX 30.1

PARKINSON'S DISEASE

Parkinson's disease (PD) is the second most common progressive neurodegenerative disorder. Patients with PD experience slowness of movement, rigidity, a low-frequency rest tremor, and difficulty with balance. These major motoric features of PD are due to the degeneration of dopamine (DA) containing neurons in the substantia nigra pars compacta (SNC) with an accompanying loss of DA and its metabolites in the striatum. PD patients may exhibit other symptoms, including cognitive dysfunction, depression, anxiety, autonomic problems, and disturbances of sleep, which may be due to the degeneration of non-DA-containing neurons. PD is characterized pathologically by the cytoplasmic accumulation of aggregated proteins with a halo of radiating fibrils and a less defined core known as Lewy bodies. Measures of increased oxidative stress are also seen, including glutathione depletion, iron deposition, increased markers of lipid peroxidation, oxidative DNA damage and protein oxidation, and decreased expression and activity of mitochondrial complex 1 in the SNC. This mitochondrial complex 1 defect is specific to the SNC, as it has not been observed in other neurodegenerative diseases. Thus, in sporadic PD, oxidative stress and mitochondrial dysfunction appear to play prominent roles in the death of DA neurons. The majority of PD appears to be sporadic in nature, but an estimated 10% of cases are familial, with specific genetic defects. The identification of these rare genes and their functions has provided tremendous insight into the pathogenesis of PD and opened up new areas of investigation.

Several genes have been clearly linked to PD. The best studied are α-synuclein and parkin. α-Synuclein is a presynaptic protein and a primary component of Lewy bodies. Very rare missense mutations at A53T and A30P cause autosomal-dominant PD. How derangements in α-synuclein lead to dysfunction and death of neurons is not known. However, aggregation and fibrillization are thought to play a central role, perhaps leading to a dysfunction in protein handling by inhibiting the proteasome. The second "PD gene" parkin is an E3 ligase for the ubiquitination of proteins. Mutations in parkin include exonic deletions, insertions, and several missense mutations and result in autosomal recessive PD. Parkin mutations are now considered to be one of the major causes of familial PD. Substrates for parkin include the synaptic vesicle–associated protein CDCrel-1, which may regulate synaptic vesicle release in the nervous system. Whether CDCrel-1 is involved in the release of dopamine is not yet known. Synphilin-1 is a protein of unknown function that was identified as an α-synuclein interacting protein. It is a synaptic vesicle-enriched protein and is present in Lewy bodies. Synphilin-1 is a target of parkin, and expression of synphilin-1 with α-synuclein results in the formation of Lewy body–like aggregates that are ubiquitinated in the presence of parkin. Parkin is upregulated by unfolded protein stress, and expression of parkin suppresses unfolded protein stress-induced toxicity. The unfolded putative G-protein-coupled transmembrane receptor, the parkin-associated, endothelial-like receptor (Pael-R), is a parkin substrate. When over-expressed, Pael-R becomes unfolded, insoluble, and causes unfolded protein-induced cell death. Parkin ubiquitinates Pael-R, and coexpression of parkin results in protection against Pael-R-induced cell toxicity. Pael-R accumulates in the brains of autosomal-recessive Parkinson's disease patients and thus may be an important parkin substrate.

Ultimately, it would be of interest to link all the multiple genes and the sporadic causes of PD into a common pathogenic biochemical pathway. Derangements in protein handling seem to be central in the pathogenesis of familial PD. Oxidative stress seems to play a prominent role in sporadic PD, and oxidative stress leads to synuclein aggregation and/or proteasomal dysfunction. Data suggest that there may be selective derangements in the proteasomal system in the substantia nigra of sporadic PD. Thus, protein mishandling may be central to the pathogenesis of PD.

In contrast to most other neurodegenerative disorders, there is effective temporary symptomatic treatment for PD consisting of DA replacement with levodopa or DA agonists and adjunctive medications or surgical approaches. Because the neurodegeneration in PD is progressive and there is no proven preventative, restorative, or regenerative therapy, patients eventually become quite disabled. Ultimately, to affect a true cure, the underlying mechanisms of neuronal cell death need to be understood, strategies for enhancing neuronal survival and regrowth need to be developed, and consideration needs to be made toward the replacement of cells lost during the degenerative process. If these goals can be obtained, a full recovery could be achieved.

Valina L. Dawson and Ted M. Dawson

Adapted from Dawson T. M. (2000). New animal models for parkinson's disease. *Cell, 101*(2), 115–118.

Adapted from Zhang Y., Dawson V. L., & Dawson T. M. Oxidative stress and genetics in the pathogenesis of parkinson's disease. *Neurobiology of Disease, 7*(4), 240–250.

The DA input to the striatum terminates largely on the shafts of the dendritic spines of medium spiny neurons (Fig. 30.4). The location of dopaminergic terminals puts them in a position to modulate transmission from the cerebral cortex to the striatum. The action of DA on striatal neurons depends on the type of DA receptor involved. Five types of G-protein-coupled DA receptors have been described (D1–D5). These have been grouped into two families based on their response to agonists. The D1 family includes D1 and D5 receptors, and the D2 family includes D2, D3, and D4 receptors. D1 receptors stimulate adenylate cyclase activity and potentiate the effect of cortical input to striatal neurons. D2 receptors inhibit adenylate cyclase activity and decrease the effect of cortical input to striatal neurons (see Chapters 5, 7, and 8).

Medium spiny neurons contain the inhibitory neurotransmitter GABA. In addition, medium spiny neurons have peptide neurotransmitters that are co-localized with GABA. Based on both the type of neurotransmitters and the type of DA receptor they contain, medium spiny neurons can be divided into two populations. One population contains GABA, dynorphin, and substance P and primarily expresses D1 receptors. These neurons send axons to GPi and to SNpr and constitute the "direct pathway" (Fig. 30.2). The predominant effect of these neurons is inhibition of their targets. The second population contains GABA and enkephalin and primarily expresses D2 receptors. These inhibitory neurons project to GPe and form the first limb of the "indirect pathway" (Fig. 30.2). The two populations of medium spiny neurons are morphologically indistinguishable and are not segregated topographically within the striatum (Fig. 30.7). This comingling of neurons suggests that they receive similar input and thus they may convey similar information to their respective targets. However, the different targets and transmitter types of the two populations provides the substrate for functional segregation within the striatum.

Although there are no apparent regional differences in the striatum based on cell morphology, an intricate internal organization has been revealed with special stains. When the striatum is stained for acetyl-cholinesterase (AChE), there is a patchy distribution of lightly staining regions within more heavily stained regions. The AChE-poor patches have been called *striosomes* and the AChE-rich areas have been called the extrastriosomal *matrix*. The matrix forms the bulk of the striatal volume and receives input from most areas of the cerebral cortex. Within the matrix are clusters of neurons with similar inputs that have been termed *matrisomes*. The bulk of the output from cells in the matrix goes to both segments of the GP and to SNpr. The striosomes receive input from the prefrontal cortex and send output to SNpc. Immuno-histochemical techniques have demonstrated that several chemicals, including substance P, dynorphin, and enkephalin, have a patchy distribution that may be partly or wholly in register with the striosomes. The striosome-matrix organization provides the substrate for another level of functional segregation within the striatum.

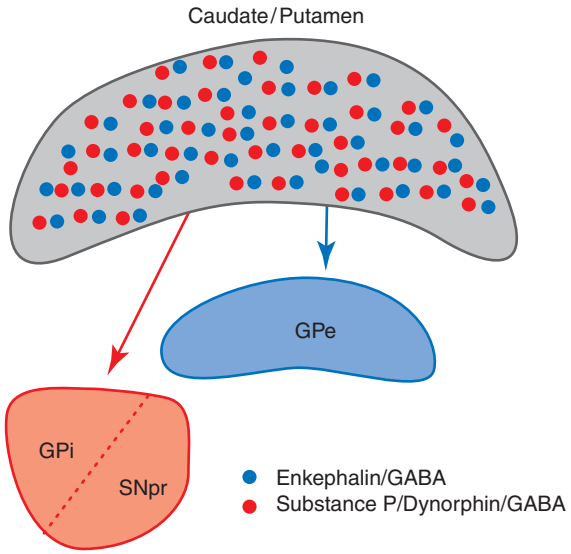


FIGURE 30.7 The two chemically different populations of striatal medium spiny neurons are intermixed. One population (blue) projects to the globus pallidus external segment (GPe) and contains GABA and enkephalin. The other population (red) projects to the globus pallidus internal segment (GPi) and substantia nigra pars reticulata (SNpr) and contains GABA, dynorphin, and substance P (red).

The Subthalamic Nucleus Receives Inputs from the Frontal Lobe

The STN is located at the junction of the diencephalon and midbrain, ventral to the thalamus and rostral and lateral to the red nucleus. Embryologically, it develops from the lateral hypothalamic cell column. The STN receives an excitatory, glutamatergic input from the frontal cortex with large contributions from motor areas, including the primary motor cortex, premotor and supplementary motor cortex, and frontal eye fields. The STN also receives an inhibitory GABA input from GPe. The output from the STN is excitatory and glutamatergic. STN projects to GPi and SNpr as well, projecting back to GPe.

Although the STN receives input from the neocortex and projects to both segments of GP and to SNpr, it is different from striatum in several ways. Unlike striatum, the cortical input to the STN is predominately from the frontal lobe. The output from STN is excitatory, whereas the output from striatum is inhibitory. Of the two routes from the cortex to GPi, GPe, and SNpr, the excitatory route through STN (5–8 ms) is faster than the inhibitory route through striatum (15–20 ms).

Globus Pallidus, Internal Segment and Substantia Nigra, and Pars Reticulata Are the Basal Ganglia Output Nuclei

GP lies medial to the putamen and rostral to the hypothalamus. Embryologically, it arises from the lateral hypothalamic cell column. In primates, GP is separated into an internal segment (GPi) and an external segment (GPe) by a fiber tract called the internal medullary lamina. In rodents and carnivores, the internal segment lies within the internal capsule and is called the entopeduncular nucleus.

The GPi is composed primarily of large neurons that project outside of the basal ganglia. The dendritic trees of GPi neurons radiate from the cell body in a disc-like distribution and are oriented so that the faces of the disc are perpendicular to incoming axons from the striatum. The dendrites of an individual GPi cell can span up to 1 mm in diameter, and therefore GPi neurons have the potential to receive a large number of converging inputs.

The principal inputs to GPi are from striatum and the STN. These two types of inputs differ in the neurotransmitters used, in their physiological action, and in their pattern of termination. As noted earlier, neurons projecting from striatum to GPi contain GABA, substance P, and dynorphin and are inhibitory. Each axon from the striatum enters GPi and sparsely contacts several neurons in passing before ensheathing a single neuron with a dense termination (Fig. 30.8). The excitatory, glutamate projection from STN to GPi is highly divergent such that each axon from the STN ensheathes many GPi neurons (Fig. 30.8). The cortico-STN-GPi pathway has been called the "hyperdirect" pathway. Thus, the striatal and STN inputs to GPi form a pattern of slow, focused, convergent inhibition from striatum ("direct pathway") and fast, widespread, divergent excitation from STN ("hyperdirect pathway") (Fig. 30.2).

The output from GPi is inhibitory and uses GABA as its neurotransmitter. The majority (70%) of the GPi output is sent via collaterals to both the thalamus and the brainstem. In the thalamus, axons from GPi terminate in the oral part of the ventrolateral nucleus (VLo) and in the principal part of the ventral anterior nucleus (VApc). In turn, these thalamic nuclei project to the motor, premotor, supplementary motor, and prefrontal cortex. They also project back to striatum in a potential feedback loop. An individual GPi neuron sends output via the thalamus to just one area of the cortex (Hoover & Strick, 1993). Thus, just as there are parallel inputs from the cortex to striatum, there appear to be functionally parallel outputs from GPi.

Collaterals of the axons projecting from GPi to the thalamus project to an area at the junction of the midbrain and pons near the pedunculopontine nucleus, which has been termed the "midbrain extrapyramidal area." The midbrain extrapyramidal area projects in turn to the reticulospinal motor system. Other GPi neurons (20%) project to intralaminar nuclei of the thalamus or to the lateral habenula. The specific role of these projections is not known.

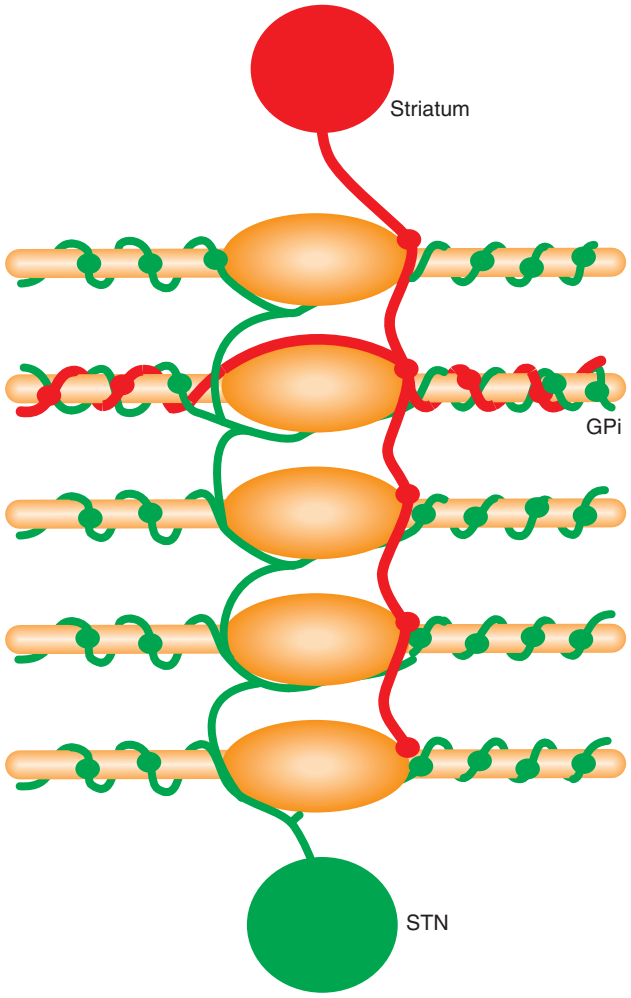


FIGURE 30.8 The two primary inputs to globus pallidus internal segment (GPi) have different patterns of termination. Axons from the subthalamic nucleus (green) are excitatory and terminate extensively on multiple GPi neurons. Axons from the striatum (red) contact several GPi neurons weakly in passing before terminating densely on a single neuron.

Like the GP, the SN is divided into two segments. One is sparsely cellular and is called the pars reticulata (SNpr), and the other is a densely cellular region called the pars compacta (SNpc). SNpr is similar to GPi in many ways, including cell size, histochemistry, and connectional anatomy. The pars compacta contains DA cells and is not a principal output nucleus of the basal ganglia.

Like GPi, the SNpr contains large neurons that project outside of the basal ganglia. Like GPi, SNpr receives inhibitory inputs from striatum that contain GABA, SP, and dynorphin and excitatory inputs from STN that contain glutamate and provides inhibitory, GABAergic outputs.

SNpr outputs are to the medial part of the ventro-lateral thalamus (VLm) and the magnocellular part of the ventral anterior thalamus (VAmc). These thalamic areas in turn project to the premotor and prefrontal cortex. Like the GPi output, SNpr sends collaterals to the midbrain extrapyramidal area and also has a projection to intralaminar nuclei of the thalamus. The primary difference in the outputs of GPi and SNpr is that the lateral portion of SNpr sends an inhibitory projection to the superior colliculus and to the paralaminar part of the medial dorsal thalamus (MDpl). MDpl projects in turn to the frontal eye fields. This lateral portion of SNpr is connected with cortical and brainstem areas which control eye movements (see Chapter 31).

GPe Is Connected to Several Other Basal Ganglia Nuclei

The GPe can be considered to be an intrinsic nucleus of the basal ganglia, as can the SNpc, which is considered in the next section. Both GPe and SNpc receive the bulk of their input from and send the bulk of their output to other basal ganglia nuclei.

The GPe is similar in some ways to GPi. Its inputs are an inhibitory projection from the striatum and an excitatory one from STN. The patterns of termination of the striatal and STN afferents are similar in that the striatal input is focused and convergent while the STN input is divergent (Parent & Hazrati, 1993). As noted earlier, GPe and GPi receive input from neighboring striatal neurons and thereby are likely to receive similar information. Unlike GPi, the striatal projection to GPe contains GABA and enkephalin but not substance P. However, the pathway taken by output of GPe is also quite different from that taken by the output of GPi, which undoubtedly has profound functional implications. The majority of the output of GPe projects to STN. The connections from striatum to GPe, from GPe to STN, and from STN to GPi thus are referred to as the "indirect" striatopallidal pathway to GPi (Alexander & Crutcher, 1990). In addition, there is a monosynaptic GABAergic inhibitory output from GPe directly to GPi and to SNpr and a recently described GABAergic projection back to striatum. Thus, GPe neurons are in a position to provide feedback inhibition to neurons in striatum and STN and feed-forward inhibition to neurons in GPi and SNpr. This circuitry suggests that GPe may act to oppose, limit, or focus the effect of the striatal and STN projections to GPi and SNpr, as well as focus activity in these output nuclei.

SNpc Provides DA Input to the Striatum

The SNpc is perhaps the most studied structure in the basal ganglia. The SNpc is made up of large DA-containing cells and it is these neurons that degenerate in PD (see later). DA neurons contain a substance called neuromelanin, which is a dark pigment that gives the SNpc a blackish appearance and appears to represent an oxidation product of DA. This is the basis for its name (substantia nigra = black substance). SNpc receives input from the striatum, specifically from the striosomes. This input is GABAergic and inhibitory. Other inputs to SNpc have been difficult to assess because the dendrites of SNpc and SNpr neurons overlap. The SNpc DA neurons project to all of caudate and putamen in a topographic manner. However, nigral DA neurons receive inputs from one striatal circuit and project back to the same and to adjacent circuits (Haber, Fudge, & McFarland, 2000). Thus, they appear to be in a position to modulate activity across circuits. The action of DA depends on the receptors located on target neurons, as was discussed earlier.

Summary of Basal Ganglia Circuitry

1. The striatum receives input from nearly all of the cerebral cortex such that several functionally related cortical areas project to overlapping striatal zones and an individual cortical area projects to several striatal zones. Cortical areas that are not functionally related project to separate zones of the striatum, although there may be some interaction between adjacent zones.
2. The striatum sends a focused and convergent inhibitory projection ("direct pathway") to the basal ganglia output nuclei, GPi and SNpr.
3. The STN receives input from the frontal lobe, especially from the motor, premotor, and supplementary motor cortex and from the frontal eye fields.
4. The STN sends a fast divergent excitatory projection ("hyperdirect pathway") to GPi and SNpr.
5. Reciprocal and loop-like connections among basal ganglia nuclei play a negative or positive feedback role and likely result in focusing of signals. These include the "indirect pathway" from striatum to GPe to STN to GPi and SNpr.
6. The output from the basal ganglia (GPi and SNpr) is inhibitory and projects to motor areas in the brainstem and thalamus.

The basal ganglia are sometimes viewed as a system with substantial convergence of information from the inputs to the outputs. In contrast, they have also been viewed as a system of multiple parallel functionally segregated loops through the basal ganglia, each with a separate output. In support of convergence are (1) the reduction of the number of neurons at each level from the cortex to striatum to GPi and SNpr, (2) the large number of synapses on each striatal neuron, (3) the large dendritic trees of striatal, GPi, and SNpr neurons, and (4) interaction across circuits mediated by SNpc and GPe neurons. In support of parallel segregated

loops are (1) the preserved somatotopy with separate representations of the face, arm, and leg in the cortex, striatum, and GPi/SNpr; (2) the relative preservation of topography through the basal ganglia; and (3) the finding that separate groups of GPi neurons project via the thalamus to separate motor areas of the cortex. It is likely that both views are correct with parallel organization of function circuits, but with convergence within circuits and interaction between adjacent (and functionally related) circuits.

SIGNALING IN BASAL GANGLIA

Although the anatomic organization may provide some clues as to what might be the function of basal ganglia circuits, inference of function from anatomy is speculative. One approach to studying the function of an area of the central nervous system is to record the electrical activity of individual neurons with an extracellular electrode in awake, behaving animals. Other approaches involve inferences of neuronal signaling from imaging studies of blood flow and metabolism, or of changes in gene expression. The activity of different regions can be correlated with performance of specific tasks. With electrophysiological studies, the timing of neural activity in one part can be compared to the timing of another and to the timing of the movement. This section will discuss movement-related signals in the basal ganglia that are correlated with the production or prevention of movement or with motor planning.

Striatal Output Neurons Have Low Spontaneous Activity That Increases With Movement

In awake animals, the majority of striatal neurons (80 to 90%) have low baseline discharge rates of 0.1 to 1 Hz (Fig. 30.9). These neurons project outside of the striatum and therefore are probably medium spiny neurons. In general, the activity of these neurons reflects the areas of cerebral cortex from which they receive inputs, though striatal neurons are less modality specific. Thus, neurons in areas of the putamen which receive input from the somatosensory and motor cortex have activity correlated with active and passive movement, but not with specific tactile modalities such as light touch, vibration, or joint position or with kinematic properties of movement such as amplitude, velocity, or force.

Striatal neurons related to movement have a distinct somatotopy with the face represented ventro-medially and the leg dorsolaterally. Neurons active in relation to eye movements are located in a longitudinal zone in the central part of the caudate nucleus. The striatum appears to be organized as clusters of neurons with similar discharge characteristics. These likely correspond to

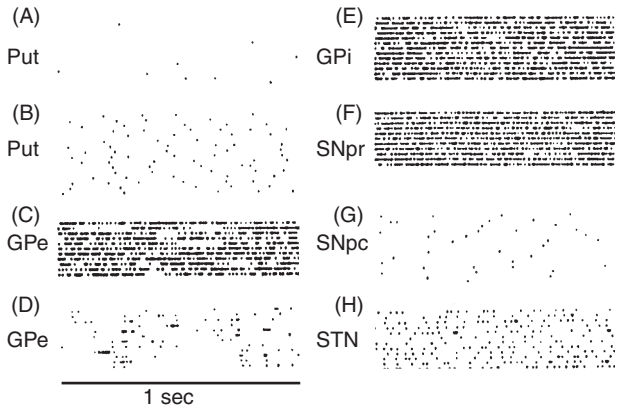


FIGURE 30.9 Representative neural discharge patterns from several basal ganglia nuclei. In the raster displays, each dot indicates the occurrence of an action potential. Each horizontal raster line represents a 1-s period of discharge. Several such periods are arranged vertically for each nucleus.

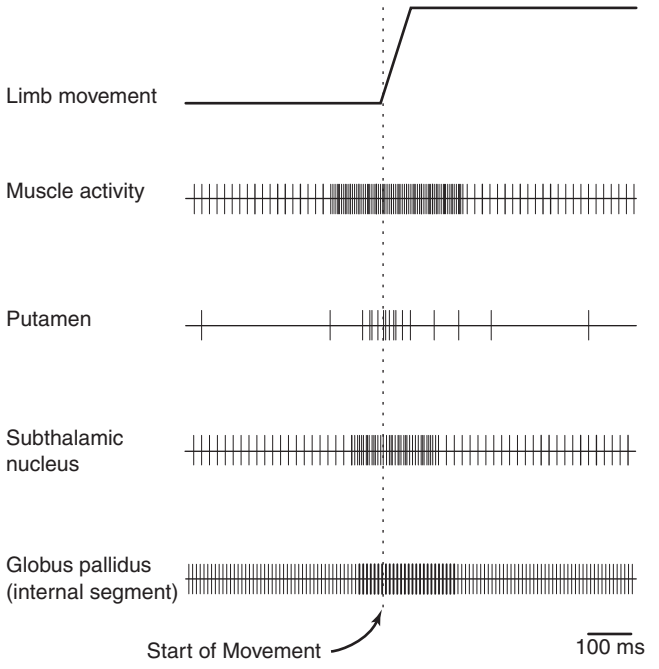


FIGURE 30.10 Schematic representation of the timing of neuronal activity in three nuclei of the basal ganglia in relation to limb movement and the muscle activity used to make the movement.

the anatomically defined matrisomes described earlier. Movement-related activity changes are typically increases above the very low baseline. On average, movement-related striatal neurons fire 20 msec prior to movement (Fig. 30.10). About half of these neurons fire in relation to the direction of movement. Some neurons fire in relation to the start of movement and others in relation to the stop. Some neurons in the putamen fire in relation to self-initiated movements, some in relation to stimulus-triggered

movements, and some in relation to both (Romo, Scarnati, & Schultz, 1992). Neurons in the anterior part of the putamen and in the caudate nucleus that receive input from the premotor or prefrontal cortex have activity during the preparation for movement. This activity is time locked to instructional cues but not to the movement itself. Some signals correlate with the instruction of whether to move ("set"). Other signals correlate with the signal to move ("go"). Although individual neurons may have signals that preferentially correlate with specific tasks or task parameters, as a whole the striatum is not exclusively involved in any one task type or parameter.

The timing of neuronal activity related to limb or eye movements is important to understanding the function of those neurons in the production of movement. In the putamen, most movement-related activity occurs relatively late, before the onset of movement but after the onset of muscle activity responsible for producing the movement. When compared to activity in the cerebral cortex, movement-related putamen neurons fire after those in the overlying motor cortex and supplementary motor area. Putamen neurons related to movement preparation also fire substantially later than cortical neurons involved in movement preparation. Caudate neurons related to eye movements fire late relative to oculomotor areas of cortex. The late timing of striatal discharge during movement and movement preparation relative to discharge in cortical areas suggests that the striatum is not involved primarily in the initiation of movement. Instead, it receives information from cortical areas that *are* responsible for movement generation and may use that information for facilitation, gating, or scaling of cortically initiated movement.

Two additional subsets of striatal neurons represent fewer than 10% of striatal neurons but are critical for basal ganglia function. One subset called the tonically active neurons (TANs) have a relatively broad action potential, are distributed throughout the striatum, and fire tonically but have no discharge correlated with movement. TANs are thought to be cholinergic interneurons which fire in relation to certain sensory stimuli that are associated with reward (Aosaki et al., 1994). It has been suggested that these neurons signal aspects of tasks that are related to learning and reinforcement. The other subset is the fast-spiking interneurons (FSIs) that are thought to be GABAergic interneurons. They have narrow waveforms, and do discharge in relation to movements. In particular, they have been shown to fire in a coordinated manner throughout striatum when animals initiate one chosen action while suppressing a highly trained alternative (Gage, Stoetzner, Wiltschko, & Berke, 2010).

Functional imaging studies using functional Magnetic Resonance Imagining (fMRI) in human subjects have shown signals in striatum related to several aspects of task performance. In addition to activity correlated with motor performance, several studies have suggested an important role of striatum in specific types of procedural memory and motor learning.

STN Has Moderate Spontaneous Activity That Increases during Movement

Neurons in the STN are tonically active with average baseline discharge rates of about 20 Hz (Fig. 30.9). They are organized somatotopically and change activity in relation to eye or limb movement. For 90% of these neurons, the change is an increase, occurring on average 50 msec prior to the movement (Fig. 30.10). The majority of movement-related neurons in STN have signals related to movement direction, but little is known about their relationship to physical parameters of movement. Functional imaging studies in human studies have demonstrated signals in STN that appear to be correlated with the stopping of ongoing movement or preventing the initiation of unwanted movements (see Box 30.8).

GPi and SNpr Have High Spontaneous Activity That Increases or Decreases after Movement Initiation

As described earlier, GPi and SNpr form the output of the basal ganglia. Neurons in these output structures fire tonically at average rates of 60 to 80 Hz (Fig. 30.9). They are organized somatotopically with the leg and arm in GPi and the face and eyes in SNpr. Due to the somatotopy, studies of limb movements have focused on GPi and studies of eye movements have focused on SNpr.

The discharge of many GPi neurons is related to the direction of limb movement, but most GPi activity is not correlated with other physical parameters of movement, including joint position, force production, movement amplitude, or movement velocity. Few GPi neurons have activity correlated with the pattern of muscle activity. Like striatum, some GPi neurons have activity related to movement preparation. During a reaching movement, 25% of GPi neurons are related to movement preparation and 50% are related to movement, but many neurons have more than one type of response. Approximately 70% of arm movement–related GPi neurons increase activity and 30% decrease activity from this tonic baseline during movement. Because the output from GPi is inhibitory, the large proportion of activity increases during movement translates to broad inhibition of thalamic and brainstem targets with more restricted facilitation during movement. GPi neurons are more likely to decrease discharge in stimulus-triggered movement and increase discharge in memory-guided movement. Thus, movement-related discharge is highly influenced by behavioral context.

The timing of movement-related GPi activity is late compared to the activation of agonist muscles. The average onset of GPi activity is after the onset of EMG but

before the onset of movement (Fig. 30.10). There is a tendency for GPi movement-related increases to occur earlier than decreases, which probably reflects the faster speed of the excitatory pathway from the cortex to GPi via STN compared with the slower inhibitory pathway from the cortex to GPi via striatum. The late timing of GPi movement-related activity suggests that the output of the basal ganglia is unlikely to initiate movement.

SNpr is the principal basal ganglia output for the control of eye movements, and most studies of SNpr activity have been in eye movement tasks (Hikosaka et al., 2000). Most saccade-related SNpr neurons fire in relation to the direction of the eye movements. Unlike GPi neurons during limb movements, virtually all saccade-related SNpr neurons decrease activity during the saccade. It is not known whether the fact that most GPi limb movement neurons increase and most SNpr saccade neurons decrease indicates a fundamental difference between eye and limb movement control or if it reflects task differences. Other differences between GPi limb movement neurons and SNpr saccade neurons have been noted. While GPi limb movement neurons have weak sensory responses, some SNpr neurons have strong responses to visual stimuli. For the visually responsive SNpr neurons, the spatial location of the stimulus seems to be the most salient feature.

Some SNpr neurons have their strongest relation to saccades made to remembered targets. If a visual target is presented and then removed and a monkey is trained to look to where the target had been, some SNpr cells will fire more in this condition as compared to when a saccade is made to a visible target. Thus, like GPi neurons, SNpr output neurons are strongly influenced by context.

GPe Has Irregular Activity That Increases or Decreases after Movement Initiation

Two types of neurons have been described in GPe based on their baseline activity patterns. Most fire at a high frequency (70 Hz) that is interrupted by long pauses. A smaller number fire at a low frequency (10 Hz on average) and have frequent spontaneous bursts of activity. Both types of neurons change activity in relation to limb movement and, for the majority, these changes are increases in activity. As has been described for GPi, GPe neurons weakly and inconsistently code for movement amplitude, velocity, muscle length, or force. Like the other structures of the basal ganglia, the activity of movement-related activity is late.

SNpc Has Low Spontaneous Activity That Does Not Change with Movement but Changes with Significant Environmental Stimuli

The activity of single neurons in the SNpc of trained animals is different from the activity of single neurons in the other basal ganglia structures. At baseline, SNpc neurons fire at a low rate (2 Hz on average). The activity of these neurons is not related to movement itself and there is no apparent somatotopy in SNpc. The neurons carry little specific information regarding sensory modality or spatial properties. The activity of SNpc neurons does change in relation to behaviorally significant events such as reward or the presentation of instructional cues. The responses to stimuli only occur if the stimulus is presented in the context of a movement task. Furthermore, the activity changes with conditioning. As an example, consider an SNpc neuron that fires in relation to an unexpected reward. If one records from this neuron over several trials while the reward is paired with a tone that precedes the reward, the SNpc neuron will gradually stop firing in relation to the reward and instead will begin to fire in relation to the tone that predicts the reward.

Thus, it appears that SNpc DA neurons can predict the occurrence of a behavioral event. In this way, SNpc neurons are similar to the TANs neurons described earlier. Remember that SNpc neurons also synapse extensively on medium spiny striatal neurons and that DA terminals are on the shafts of dendritic spines. It has been suggested that the DA input changes the sensitivity of striatal neurons to cortical inputs that terminate on the heads of dendritic spines. If so, the activity of SNpc neurons could modify the response of striatal neurons to cortical input that occurs in a specific behavioral context. Such influences may be related to the apparent ability of DA to mediate both long-term potentiation and long-term depression in striatal neurons, which in turn is likely to be an important mediator of learning and plasticity in the striatum.

Summary of Signaling in the Basal Ganglia

Several general statements about basal ganglia movement-related neuronal discharge can be made.

1. Movement-related neurons in the striatum, STN, GP, and SNpr are arranged somatotopically.
2. Medium spiny output neurons in the striatum are quiet at rest and increase during movement. Neurons in STN are tonically active and increase during movement. Recalling the anatomy, this means that GPi neurons receive a widespread, tonically active excitatory input and a focused, intermittent inhibitory input.
3. STN has increased activity when an ongoing movement is stopped suddenly or when initiation of movement is prevented.
4. Neurons in both GPe and GPi are tonically active. Most increase activity, but up to one-third decrease activity during limb movement.
5. Neurons in SNpr are also tonically active. Those that are related to saccadic eye movements decrease activity during the movement.

6. Neurons in SNpc discharge in relation to rewards and behaviorally relevant stimuli, but not to movement. They are likely to play a critical role in some types of motor learning.
7. Changes in the activity of basal ganglia occur at the onset of movement but after the muscles are already active. Thus, they are unlikely to initiate movement.
8. Movement-related activity throughout the basal ganglia is modulated by the behavioral context in which the movements are performed.

THE EFFECT OF BASAL GANGLIA DAMAGE ON BEHAVIOR

Valuable clues to the function of the basal ganglia have come from studying activity patterns during behavior. However, correlation of activity with an aspect of behavior does not necessarily mean that structure plays a causal role in that behavior. One approach to determining the precise roles of the basal ganglia in behavior is to selectively remove a specific component from an otherwise intact system. Certain human neurological diseases cause the loss of specific types of neuron in the basal ganglia. The movement disorders that result from basal ganglia damage are often dramatic. Depending on the site of the pathology, some basal ganglia diseases cause extreme slowness of movement and rigidity and others cause uncontrollable involuntary movements. Although human diseases are of great interest, they often affect more than one structure. This necessarily limits the power of functional models derived from the study of human basal ganglia diseases. In experimental animals, more selective lesions can be made with specific pharmacologic agents. More recently, mice have been genetically engineered to prevent development of specific neuron types.

Experimental Damage to the Striatum Causes Slow Voluntary Movements or Involuntary Postures and Movements

Lesions in the striatum produce variable results that depend on the location of the lesion, the lesion method, and what is measured. Many studies of unilateral striatal lesions have described only minimal deficits. If the putamen is inactivated pharmacologically with the GABA agonist muscimol unilaterally, the result is slightly slow movement of the contralateral limb. The slowed movement is associated with an increased activity of antagonist muscles. Despite slower movements after putamen lesions, reaction time is generally normal, indicating that movement initiation remains intact. Large bilateral lesions result in the paucity of movement, severe slowness of movement bilaterally, and postural abnormalities. In other studies, there appears to be little effect of bilateral electrolytic putamen lesions. The reason for this discrepancy is not clear, but it may be due to differences in lesion size.

Selective inactivation of striatal inhibitory interneurons (FSIs) results in abnormal involuntary movements (Gittis et al., 2011). The selective inactivation is thought to cause excessive firing of direct pathway striatal medium spiny neurons which, in turn, results in reduction of the inhibitory output from GPi and SNpr.

Huntington's disease (HD) is a genetically based, degenerative disease in humans that results in disabling involuntary movements (see Box 30.4). These movements are called chorea (Greek for "dance"). They are frequent, brief, sudden, random twitch-like movements that involve all parts of the body and resemble fragments of normal voluntary movement. The pathologic hallmark of HD is a marked loss of neurons in the striatum. Studies have shown that not all striatal neurons are equally affected in HD (Albin, Young, & Penney, 1989). In early adult-onset HD, the enkephalin-containing neurons that project to GPe are lost first. Clinically, this is associated with the presence of chorea. In late adult-onset and in juvenile-onset HD, both the enkephalin-containing and the substance P-containing striatal neurons that project to GPi and SNpr are lost. Clinically, this loss is associated with the presence of rigidity and dystonia (sustained, abnormal postures). Interestingly, experimental destructive lesions of the striatum in monkeys do not result in chorea. This is probably due to the nonselective destruction of both GPe-and GPi projecting striatal neurons. More selective disruption of the striatal–GPe pathway with a GABA antagonist administered into the GPe does produce chorea. The suggested mechanism for chorea is that disinhibition of GPe neurons (pharmacologically, or as a result of selective striatal cell death) causes inhibition of STN and GPi. This results in abnormal overactivity of motor cortical and brainstem mechanisms, resulting in chorea.

Damage to STN Causes Large-Scale Involuntary Movements

In monkeys and humans, lesions in the STN cause dramatic involuntary flinging movements of the contralateral arm and leg. These movements have been called hemiballismus. They resemble chorea in their brief, random, and sudden nature, but they tend to be much larger in amplitude. After a lesion in STN, the hemiballism lasts for days to weeks before gradually resolving. Animals and humans with hemiballism can still make voluntary movements and often the voluntary movements appear to be quite normal. It has been suggested that the mechanism underlying the hemiballism is a loss of excitatory input to GPi, resulting in decreased GPi activity and ultimately in the disinhibition of cortical and brainstem motor mechanisms.

Several lines of evidence support this. First, blocking the excitatory transmission from STN to GPi with a glutamate antagonist administered into GPi causes involuntary movements. Second, metabolic activity patterns as shown with 2–deoxyglucose uptake indicate decreased GPi output in hemiballism. Third, neuronal recording during stereotaxic surgical treatment of movement disorders has shown decreased GPi activity in patients with chorea or hemiballism. However, it is likely that chorea or hemiballism is due to abnormal patterns of basal ganglia output signals and not just tonically decreased activity. Thus, if hemiballism or chorea is produced by an STN lesion in monkeys, a second lesion placed in GPi abolishes the involuntary movements. Furthermore, in monkeys with STN lesions, the level of GPi activity is reduced even after the involuntary movements resolve. Finally, destruction of GPi in humans with hemiballism or chorea abolishes the involuntary movements and experimental lesions of GPi in monkeys do not cause chorea. Therefore, it seems likely that chorea and hemiballism reflect an abnormal fluctuating or bursting output from the basal ganglia rather than a simple reduction of output.

In rodent studies, involuntary movements are uncommon but STN lesions cause animals to respond prematurely to cues and impair the ability to prevent unwanted movements. Similar difficulties with impulse control occur in people with Parkinson disease who have been treated with Deep Brain Stimulation of STN (see Box 30.5).

Damage to GP Causes Slow Voluntary Movements and Involuntary Postures and Does Not Delay Movement Initiation

Lesions of the GPi result in slowing of movement but normal movement initiation. Because of their close proximity to each other, it is difficult to lesion GPi without including part of GPe or vice versa. Therefore, caution must be used in interpreting results of GP lesions. However, in most cases, combined GPe and GPi lesions have a similar effect to the lesion of GPi alone. Unilateral lesions of GPe and GPi result in slowness of movement, with abnormal cocontraction of agonist and antagonist muscles, but normal initiation of movement. Bilateral lesions of GPe and GPi result in even more severely abnormal flexed postures with an apparent inability to move out of them. Electrolytic lesions producing this severe abnormality have tended to be large, involving both GPe and GPi and part of the internal capsule. Similar abnormalities are seen with carbon monoxide or carbon disulfide poisoning, which results in neuron death in GPi and SNpr.

Lesions restricted to GPi result in slowness of movement of the contralateral limbs. After GPi lesions, reaction time is normal, which indicates that mechanisms involved in the initiation of movement are intact. In some studies, the slowness of movement is accompanied by a cocontraction of agonist and antagonist muscles, which results in abnormal postures similar to those that occur in dystonia, a basal ganglia movement disorder. In other studies, the slowing may depend on the context of the task. Monkeys with GPi lesions are more impaired when movements are made by turning off active muscles than when movements are made by further turning on active muscles (Fig. 30.11). Thus, lesions that remove the inhibitory output of the basal ganglia appear to interfere with the ability to turn off unwanted muscle activity.

Damage to SNpr Causes Involuntary Eye Movements

Due to its proximity to SNpc, it is difficult to produce electrolytic lesions exclusively in SNpr. However, with the use of the GABA agonist muscimol, it has been possible to inactivate neurons focally in SNpr. Injection of muscimol into the lateral SNpr inactivates neurons that are normally involved in saccadic eye movements. Inactivation in this area results in an inability to maintain visual fixation because of involuntary saccades (Fig. 30.12). The inability to suppress involuntary saccades appears to result from disinhibition of the superior colliculus. Injection of the GABA antagonist bicuculline into the superior colliculus mimics the effects of muscimol in SNpr. Thus, just as GPi inactivation results in abnormal excess limb and trunk muscle activity, SNpr inactivation results in abnormal excess eye movements.

Damage to SNpc Causes Symptoms of Parkinson's Disease

Lesions of SNpc are of particular interest because the DA neurons in this nucleus die in Parkinson's disease (PD). PD is a neurological disease that affects older adults (see Box 30.1). The main symptoms of PD are (1) tremor at rest that decreases during movement, (2) slowness of movement (bradykinesia), (3) paucity of movement (akinesia), (4) muscular rigidity, and (5) unstable posture. The primary pathology in PD is a progressive degeneration of neurons in the SNpc. Additionally, there is some degree of loss of DA neurons in the ventral tegmental area and of norepinephrine neurons in the locus coeruleus. Although PD has been recognized for over a century, it has only been since the early 1980s that a complete animal model has been available. As noted earlier, the interdigitation of SNpc and SNpr has made selective electrolytic lesions of SNpc or SNpr difficult to produce. In the 1960s, it was found that 6-hydroxydopamine (6-OHDA) produces specific lesions of catecholamine neurons. Therefore, injection of 6-OHDA into the rat SN bilaterally produces

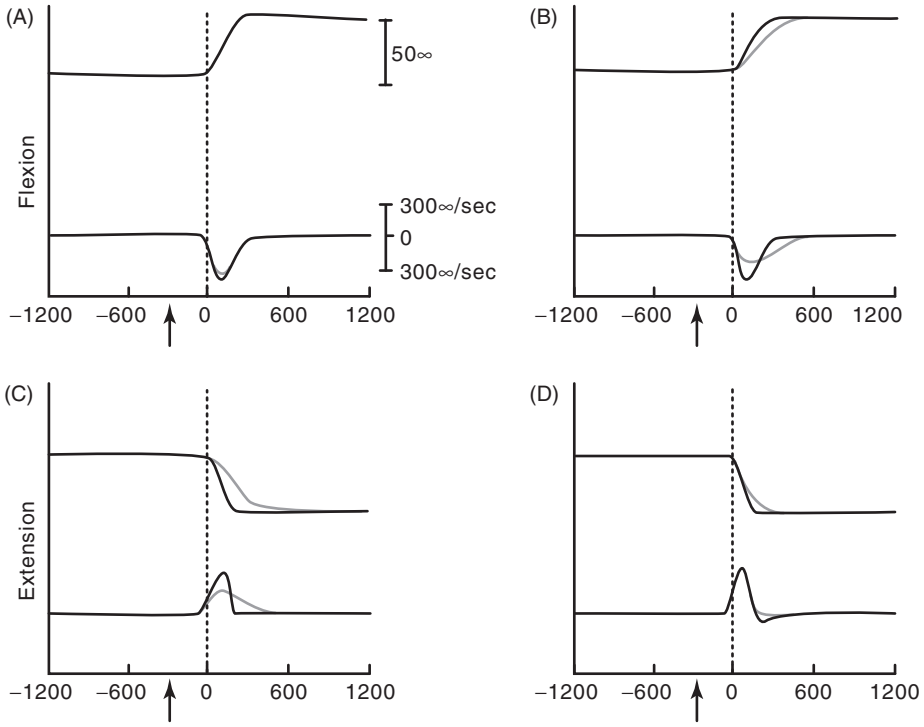


FIGURE 30.11 Wrist position and velocity in visually guided wrist movements before (black traces) and after (blue traces) a lesion of the globus pallidus internal segment. In each graph, the top traces represent wrist position, the lower traces represent wrist velocity. (A) Flexion with the flexor muscles loaded (movement made by further activating the loaded muscles). (B) Flexion with extensor muscles loaded (movement made by turning off the loaded muscles). (C) Extension with flexor muscles loaded (movement made by turning off the loaded muscles). (D) Extension with extensor muscles loaded (movement made by further activating the loaded muscles). After the lesion, the peak velocity was slower when the movements were made by decreasing the activity of the loaded muscle than when made by increasing the activity of the loaded muscle. *From Mink and Thach (1991).*

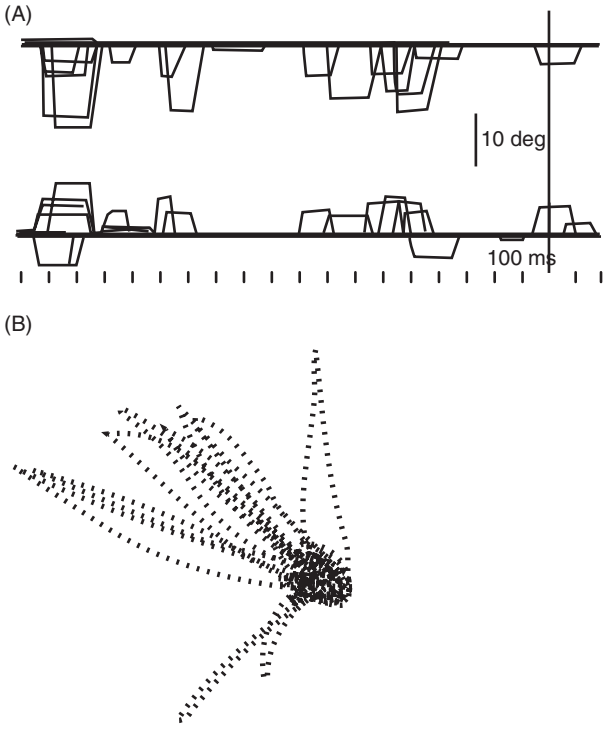


FIGURE 30.12 After inactivation of substantia nigra pars reticulata, monkeys are unable to maintain fixation of gaze because of involuntary contraction of the eye muscles. (A) The top line represents vertical eye position, and the bottom line represents horizontal eye position during attempted visual fixation. (B) Lines represent the trajectory of the involuntary eye movement when the monkey was instructed to maintain its gaze in the center dot. *From Hikosaka and Wurtz (1985).*

a selective loss of SNpc DA neurons. Lesions made in this way result in some of the abnormalities of PD—namely, the slowness of movement and rigidity, but they did not produce tremor. In the 1980s, a street drug contaminant called MPTP (see Box 30.2) was found to produce all of the symptoms of PD in humans and in certain species of monkeys. The MPTP monkey has proven an extremely valuable model with which to explore this disease.

Despite extensive investigation, the fundamental mechanism of the tremor in PD is not known. Some evidence suggests that it results from abnormal bursting of neurons in the thalamus. The slowness of movement in PD has been associated with a reduced magnitude and duration of muscle activity during movement. In monkeys that have been given 6-OHDA or MPTP, the activity of motor cortex neurons during arm movements is also reduced compared with normal monkeys. In the MPTP monkey model of PD, some animals have abnormally increased activity of STN and GPi and decreased activity of GPi. GPi neurons in the MPTP monkey have abnormal bursting activity and abnormally increased responses to somatosensory stimuli. It has been suggested that the increased activity of GPi causes an excess inhibition of motor mechanisms in the cortex and brainstem and that this excess inhibition causes movements to be slow (Albin et al., 1989). However, it has been argued that the abnormal patterns of GPi activity may be more important than the more modest increase in tonic activity.

The rigidity of PD has been attributed in part to hyperactivity of the transcortical stretch reflex. Normally, the

BOX 30.2

THE MPTP STORY

Until the early 1980s, the quest for a complete animal model of Parkinson's disease (PD) was largely unsuccessful. Although some of the abnormalities of PD could be caused by electrolytic lesion of SNpc or by injecting the neurotoxin 6-hydroxydopamine into SNpc, neither of these methods produced the full syndrome of PD. Then, in 1982 an unfortunate, but fortuitous, accident happened. Four young drug users in northern California developed symptoms of PD. Because PD is highly unusual in young adults, the neurologists caring for these patients began a search for the cause (Langston, Ballard, Tetrud, & Irwin, 1983). They discovered that each of the patients had recently obtained a "new" synthetic heroin. This "new heroin" contained an analog of the narcotic meperidine, 1-methyl-4-phenyl-proprionoxy-piperidine (MPPP), and a contaminant, 1-methyl-4-phenyl-1,2,5,6-tetrahydropyridine (MPTP). It turned out that MPTP was the agent responsible for the parkinsonism. MPTP is oxidized in the brain to MPP^+ by monoamine oxidase. MPP^+ is taken up by DA neurons where it inhibits oxidative metabolism in the mitochondria and ultimately leads to cell death.

Subsequently, it has been found that MPTP produces a Parkinson's-like syndrome in several species of monkeys. Unlike previous models, monkeys given MPTP had not only slowness of movement, paucity of movement, and rigidity, but also tremor. The Parkinson's-like syndrome in monkeys given MPTP is associated with a nearly complete degeneration of DA neurons in SNpc and the ventral tegmental area and variable degeneration in the locus coeruleus. MPTP monkeys improve when given the DA precursor L-DOPA and have side effects of chorea when they are given too much L-DOPA, similar to what happens in people with PD. Hence, by behavioral, pharmacological, and pathological measures, the MPTP monkey is an excellent model of human PD. It is currently used to study the pathophysiology and pharmacology of PD and it has lead to new ideas about possible causes of PD. One such idea is that environmental toxins may play a role in the etiology of PD. In this regard, it is important to note that paraquat and rotenone, two compounds that are used commonly as pesticides and/or herbicides, both have MPTP-like structures and have also been shown to damage DA neurons.

Jonathan W. Mink

Reference

Langston, J. W., Ballard, P., Tetrud, J. W., & Irwin, I. (1983). Chronic parkinsonism in humans due to a product of meperidine-analog synthesis. *Science, 219,* 979–980.

transcortical stretch reflex is active to resist displacement from an actively held posture. It is inhibited when subjects are instructed not to resist the displacement. People with PD have abnormally increased transcortical stretch reflexes and are unable to suppress them in response to instruction. The inability to inhibit long-loop reflexes may also account for the postural instability of PD. Patients with PD have an inappropriate cocontraction of leg and back muscles in response to perturbation from an upright stance. When the same subjects perturbed from a sitting position, they are not able to inhibit the postural reflexes that were active during stance. This suggests that the mechanism of rigidity and postural instability may be similar and that they reflect an inability to suppress unwanted reflex activity.

Paradoxically, it is known that lesions of GPi or STN can ameliorate many or all of the signs and symptoms of PD without producing the abnormalities associated with lesions in these areas in normal animals or humans. This is particularly paradoxical for GPi lesions that can produce the signs of PD (flexed posture, slow movement, rigidity) in normal monkeys and in some rare human diseases, but can reduce or eliminate those signs in monkeys and humans with preexisting lesions of the DA system. Why this is the case is not known, but it may relate to long-term changes in basal ganglia physiology that result from chronic DA depletion. Chronic, high-frequency stimulation of the STN or GPi has been found to be effective for the treatment of symptoms in advanced PD. This has been called deep brain stimulation (DBS; see Box 30.5). The mechanisms by which DBS works are not well understood, but it is thought that high-frequency stimulation disrupts neural signaling and thus mimics a lesion in the stimulated area. The growth of neurosurgical treatment of PD is a direct result of knowledge about basal ganglia circuitry.

Summary

1. Damage to any basal ganglia structure may cause slowness of voluntary movement, involuntary movements, involuntary postures, or a combination of these.
2. Damage to the striatum causes voluntary movements to be slow and may produce involuntary movements or postures depending on the mechanism of damage.

3. Damage to the STN causes large amplitude involuntary limb movements.
4. Damage to the GP causes slowness of movement, abnormal postures, and difficulty relaxing muscles, but does not delay movement initiation.
5. Damage to SNpr causes abnormal eye movements, but does not delay the initiation of eye movements.
6. Damage to SNpc causes tremor at rest, slowness of movement, rigidity, and postural instability, which are the main features of PD.

PRINCIPLES OF BASAL GANGLIA OPERATION FOR MOTOR CONTROL

How do the basal ganglia participate in motor control? Several hypotheses have been advanced over the past century. A growing consensus suggests that models based on human disease states may not be sufficient to explain normal basal ganglia function. Nonetheless, the models based on human disease have been useful in developing new treatments for movement disorders and for the development of testable hypotheses relating to basal ganglia function.

An old model of basal ganglia function is that the basal ganglia initiate movement. This model was based in large part on the manifestation of basal ganglia diseases. The paucity and slowness of movement in PD were attributed to an inability to initiate movements, whereas the involuntary movements of chorea and hemiballism were attributed to a release of normal motor systems from basal ganglia control. This model gained support from the fact that much of the output from the basal ganglia goes to parts of the thalamus that project to the premotor and motor cortex. It was argued that motor programs are stored in the basal ganglia and are called up and sent to the motor cortex for execution. This model is no longer widely accepted because it is now apparent that basal ganglia are active relatively late in relation to movement and to the activation of those brain mechanisms that are known to be involved in initiation. Furthermore, lesions of basal ganglia output do not delay the initiation of movement.

If basal ganglia do not initiate movement, what do they do? We consider three current hypotheses. One hypothesis states that although basal ganglia do not initiate movement, they contribute to the automatic execution of movement sequences. This hypothesis suggests that other mechanisms initiate the first component in a sequence, but that basal ganglia contain the programs for completion of the sequence. The second hypothesis states that the basal ganglia circuitry is made up of opposing parallel pathways that adjust the magnitude of the inhibitory GPi output in order to increase or decrease movement. According to this hypothesis, increased GPi output slows movements and decreased GPi output increases movement. The third hypothesis states that basal ganglia act to permit desired movements and to inhibit unwanted competing movements.

Do Basal Ganglia Automatically Generate Learned Movement Sequences?

A popular hypothesis states that basal ganglia are responsible for the automatic execution of learned movement sequences (Marsden, 1987). It has been pointed out that patients with PD have difficulty moving several body parts simultaneously or sequentially, and that this difficulty is more than one would expect from a simple addition of the deficits of each component of the movement. An example of an apparent problem with sequential movements in PD is the phenomenon of micrographia (or small writing). A patient begins to write a sentence with nearly normal-sized writing, but within several letters, the writing begins to get smaller so that by the end of the sentence, it may be illegible. It has been emphasized that the early components of the sequence are larger and faster than are the subsequent components. One experiment compared the performance of elbow flexion and hand grip individually or in sequence (Benecke, Rothwell, Dick, Day, & Marsden, 1986). Patients with PD performed each movement more slowly than normal subjects. However, when the movement was part of a sequence, it was slowed to an even greater degree than when it was performed separately. Another experiment involved recording the activity of GPi neurons in monkeys trained to perform two successive prompt wrist movements. It was found that some GPi neurons fired after the first component of the movement but before the second component. Proponents of the sequencing hypothesis speculate that the loss of this GPi output signal in PD is responsible for the relatively greater difficulty in producing sequential movements than in producing individual movements.

Do Basal Ganglia Produce or Prevent Movement by Using Opposing Direct and Indirect Parallel Pathways?

This hypothesis emphasizes the two major paths of information flow from the striatum to GPi and SNpr that were described earlier (Fig. 30.2). To recapitulate, one is an inhibitory "direct" pathway from striatum to GPi/SNpr and the other is a net excitatory "indirect" pathway from striatum to GPe (inhibitory), from GPe to STN (inhibitory), and from STN to GPi/SNpr (excitatory). In this hypothesis, the two pathways are in balance in such a way that increased activity in the "direct" pathway causes decreased GPi/SNpr output and increased activity in the "indirect" pathway causes increased GPi/SNpr output. By adjusting the balance, cortical targets of the basal ganglia can be

facilitated or inhibited. The hypothesis predicts that abnormally decreased output results in excessive movements (chorea) and abnormally increased output results in a decreased movement (PD). The "direct/indirect pathway" model has been useful for understanding and developing certain treatments of movement disorders, but it has several shortcomings when it comes to explaining normal basal ganglia function and the mechanism of treatments for involuntary movements. Despite the shortcomings, this has been a central hypothesis in the field of basal ganglia research for over 20 years. If the success of a model can be measured by the amount of research it stimulates, this one has been highly successful.

Do Basal Ganglia Select and Inhibit Competing Motor Patterns?

The output of the basal ganglia is inhibitory to posture and movement pattern generators in the cerebral cortex (via thalamus) and in the brainstem. The inhibitory output neurons fire tonically at high frequencies. In this hypothesis, the motor output of the basal ganglia is analogous to a brake (Mink, 1996). The hypothesis states that when a movement is initiated by a particular motor pattern generator, GPi neurons projecting to that generator decrease their discharge, thereby removing tonic inhibition and "releasing the brake" on that generator. GPi neurons projecting to other movement pattern generators increase their firing rate, thereby increasing inhibition and applying a "brake" on those generators. Thus, other postures and movements are prevented from interfering with the one selected.

How might this mechanism work (Fig. 30.13)? When one makes a voluntary movement, that movement is initiated by the prefrontal, premotor, and motor cortex and by the cerebellum. The premotor and motor cortex send a corollary signal to STN, exciting it. STN projects to GPi in a widespread pattern and excites GPi. In parallel, signals are sent from the cortex to the striatum, which inhibits GPi focally via a direct pathway. Striatum can also disinhibit GPi via two indirect pathways (striatum → GPe → GPi and striatum → GPe → STN → GPi). The indirect pathways further focus the effects of the fast excitatory cortico-STN pathway and the slower inhibitory cortico-striatal pathway to GPi. The net result is to release the "brake" from the selected voluntary movement pattern generator and to apply the "brake" to potentially competing posture-holding pattern generators (transcortical, vestibular, tonic neck, and other postural reflexes; see Chapter 28). The result is the focused selection of desired motor patterns and surround inhibition of competing patterns. Disruption of the ability to facilitate desired movements and inhibit unwanted movements results in slow voluntary movements (parkinsonism), abnormal involuntary movement (chorea, dystonia, tics), or both (Mink, 2003).

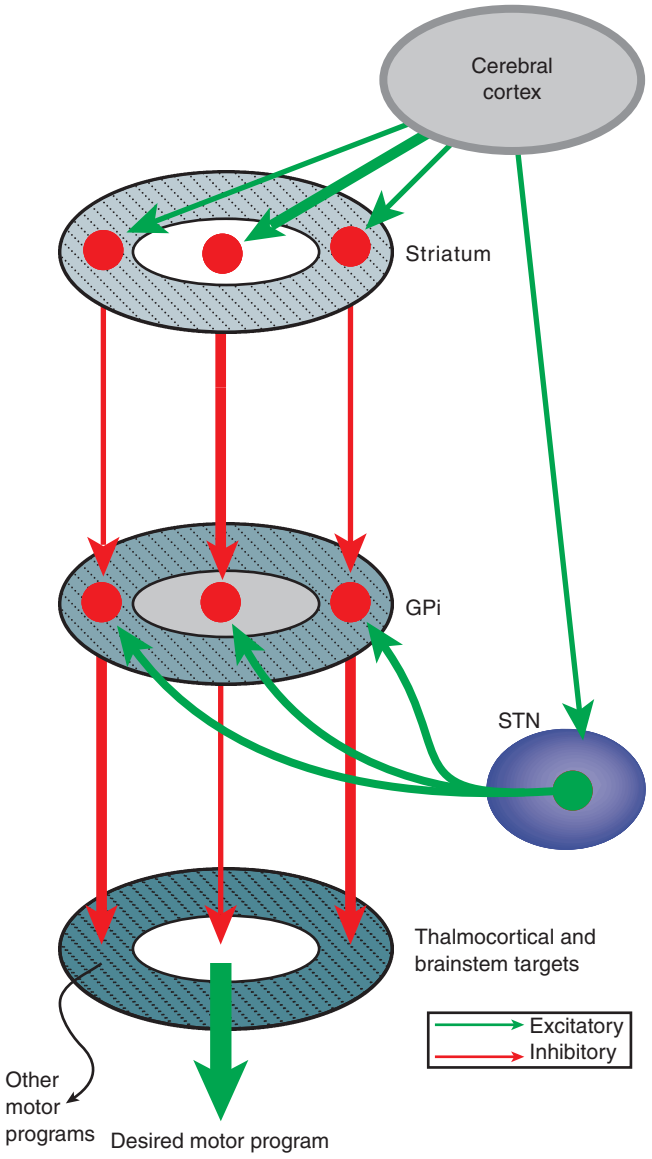


FIGURE 30.13 Relationship of proposed center-surround organization of GPi to inputs from the striatum and subthalamic nucleus. During voluntary movement, excitatory subthalamo-pallidal neurons increase the activity of the pallidal neurons in the surround territory. Inhibitory striatopallidal neurons inhibit the functional center in a focused manner. Pallidal activity changes are conveyed to the targets in the thalamus and midbrain brainstem, causing disinhibition of neurons involved in the desired motor program and inhibition of surrounding neurons involved in competing motor programs. Excitatory projections are indicated with green arrows; inhibitory projections are indicated with red arrows. The relative magnitude of activity is represented by line thickness.

BASAL GANGLIA PARTICIPATION IN NONMOTOR FUNCTIONS

Although the focus of this chapter has been on motor control, it has become increasingly clear that the basal ganglia participate in a variety of nonmotor functions. These include functions of the limbic system (see Box 30.3) and

cognitive functions. The anatomy of basal ganglia circuits has revealed outputs to all areas of frontal cortex, placing the basal ganglia in a position to influence a wide variety of behaviors. As discussed earlier, specific areas of the basal ganglia are connected preferentially with specific areas of the cerebral cortex. This provides the anatomic substrate for the localization of different functions in the basal ganglia circuits. It is known that focal lesions of the striatum tend to produce deficits similar to those seen after lesions of afferent areas of the cortex. Thus, lesions of posterior putamen cause movement deficits, lesions of inferior caudate produce deficits similar to lesions of the orbitofrontal cortex, and lesions of the dorsolateral caudate produce deficits similar to lesions of dorsolateral prefrontal cortex. The basal ganglia have been implicated in a variety of nonmotor disorders, including depression, obsessive-compulsive disorder (see Box 30.6), attention deficit hyperactivity disorder, and schizophrenia. It is becoming clear that the basic underlying principles of basal ganglia function are similar. Thus, the intrinsic circuitry is the same for cognitive and motor parts of the basal ganglia, the basic neurophysiology appears to be the same, and the effects of lesions are analogous. For example, chorea is characterized by excessive involuntary movements, and obsessive-compulsive disorder is characterized by excessive involuntary thoughts and complex behaviors. The hypothesis of focused selection and surround inhibition has been applied to non-motor basal ganglia functions (Redgrave, Prescott, & Gurney, 1999) and data support each hypothesis from cognitive science and psychiatric research.

Increasing evidence points to a role for the basal ganglia in procedural learning. Much of the research has focused on the learning of tasks or of sequential behavior. There is evidence for a role of the basal ganglia in procedural learning that leads to the formation of habits and the performance of behavioral routines once they are learned (Jog, Kubota, Connally, Hillegart, & Graybiel, 1999). As described earlier, TANs and SNpc DA neurons fire in relation to behaviorally significant events. The activity patterns of these neurons change as the task becomes learned and when novel stimuli or events are introduced. There is also growing evidence for DA-mediated long-term potentiation and long-term depression in striatal neurons that is likely to play an important role in learning. It has been shown that non-TAN striatal neurons also change activity in relation to learning. In rats performing a T-maze task, striatal neurons changed activity patterns as the task became learned and more automatic (Jog et al., 1999). In monkeys performing a discrimination learning task, striatal neurons changed activity as the animal learned new associations between stimuli and reward (Tremblay, Hollerman, & Schultz, 1998). Functional imaging studies have also shown basal ganglia activity correlated with the learning of new tasks. Striatal lesions or focal striatal DA depletion impairs the learning of new movement sequences, and procedural learning has been shown to be impaired in people with PD. These findings together support a role for the basal ganglia in certain types of procedural learning. However, other brain structures have also been implicated in procedural learning, including the cerebellum (see Chapter 31). Just as the cerebellum and basal ganglia play different roles in motor control, they are likely to play different, but complementary, roles in procedural learning. The exact nature of those different roles is the subject of current research.

BOX 30.3

VENTRAL SYSTEM OF THE BASAL GANGLIA

There are nuclei in the brain that historically have not been included with the basal ganglia, but that are analogous to the traditional components. These lie ventral to the neostriatum and GP and are called the ventral striatum (nucleus accumbens and olfactory tubercle) and the ventral pallidum. The ventral striatum receives input from limbic and olfactory areas of the cortex, including the amygdala and hippocampus. Like the neostriatum, the ventral striatum receives a *DA* input, but it is from the ventral tegmental area (VTA), which lies medial to the SN. The ventral striatum sends a projection back to VTA and to the adjacent SNpr. The ventral pallidum is analogous to both GPe and GPi. It receives input from the ventral striatum and possibly from STN. Unlike the GP, the ventral pallidum receives direct input from the amygdala. The output of the ventral pallidum projects to the dorsomedial nucleus of the thalamus (DM) and from there to limbic areas of the cortex. By virtue of its inputs and outputs, the ventral system is closely linked to the limbic system. This linkage has led to the suggestion that the ventral system is involved to some degree in motivation and emotion (see Chapter 41). The exact nature of this role is not known, but it may be analogous to the motor role of the basal ganglia, with the inhibitory output of the ventral pallidum acting to suppress or select potentially competing limbic mechanisms.

Lesions of the ventral pallidum lead to an inability to inhibit incorrect responses in an odor discrimination task, supporting the hypothesis.

Jonathan W. Mink

BOX 30.4

HUNTINGTON'S DISEASE

Huntington's disease (HD) is an inherited progressive neurodegenerative disorder that affects about 1 in 10,000 people. Symptoms include abnormal movements (comprising both involuntary movements termed chorea or dystonia and motor incoordination), cognitive difficulties, and emotional difficulties, including depression, apathy, and irritability. Onset is usually in midlife, but can range between childhood and old age. HD is generally fatal within 15 to 20 years after onset. The disease is caused by an expanded CAG repeat coding for polyglutamine in the HD gene product, termed the "huntingtin" protein. CAG repeat lengths between about 10 and 25 are normal, whereas those above 36 cause HD. Within the expanded range, the longer the repeat, the earlier the age of onset. The normal function of huntingtin is poorly understood, but it may be involved with cytoskeletal function, or cellular transport.

Pathologically, HD is characterized by selective neuronal vulnerability. The caudate and putamen of the corpus striatum are most affected in early stages, but as the disease progresses, other areas of the brain also become affected. Within the striatum, medium spiny GABA neurons are severely affected, with up to 95% loss in advanced cases, whereas large interneurons are relatively spared. In addition, there are intranuclear inclusion bodies and perinuclear and neuritic aggregates of huntingtin in HD neurons.

The pathogenesis of HD is still incompletely understood and represents a challenge for neuroscience. The huntingtin protein is widely expressed throughout the brain and is expressed at lower levels in other regions of the body as well. HD, like other neurodegenerative diseases, involves abnormal protein folding and aggregation. The inclusions themselves appear not to be directly toxic, but they likely represent a late stage of an abnormal process, whose earlier steps are pathogenic. The study of the disease has been facilitated by the generation of cell models, invertebrate models, and mouse models. Chaperone proteins can ameliorate the pathology, perhaps by refolding abnormally folded huntingtin. Huntingtin may normally be degraded by intracellular proteolytic machinery, termed the proteasome, and inhibition of the proteosome may be one of the toxic effects of the abnormally expanded polyglutamine.

A number of lines of evidence have pointed to a possible role for abnormal gene transcription in HD. Several studies have suggested that targeting of huntingtin to the nucleus enhances toxicity. Huntingtin may normally have a role in regulation of gene transcription, although this is poorly understood. Expression array studies show that mouse and cell models have an alteration of normal gene expression patterns. Abnormal intereactions with coactivator and corepressor complexes may contribute.

Whatever the initial mechanisms, symptoms appear to be caused by both cell dysfunction and cell death, with cell death correlating best with functional disability. Thus, preventing cell death is a major goal of experimental therapeutics in HD and other neurodegenerative disorders. Cell death may partly involve apoptotic mechanisms with activation of caspases and may also involve mechanisms such as excitotoxicity, metabolic, or free radical stress. Therapeutic strategies under investigation include protection of neurons with neurotrophins, altering abnormal gene transcription patterns, and reducing cell death with inhibitors of caspase activation or blockers of excitotoxicity.

Christopher A. Ross

BOX 30.5

DEEP BRAIN STIMULATION

Ablative neurosurgery has been used for treating movement disorders, epilepsy, and pain for over 80 years. In ablative procedures, a lesion is placed in the brain in order to disrupt a dysfunctional circuit. Ablative neurosurgery is still used, but increasingly deep brain stimulation (DBS) is replacing ablations to treat disorders associated with neural circuit dysfunction. This is particularly true for movement disorders related to basal ganglia dysfunction.

The idea of treating neurologic disorders with chronic electrical stimulation emerged in the 1960s. By the 1990s, technological advances provided the possibility of a system that could be implanted with no external components that could provide continuous stimulation for years without battery replacement. DBS involves stereotaxic surgical placement of electrodes into specific brain targets, fixing the electrodes to the skull, passing the electrode wires under the skin and ultimately connecting to a pacemaker-like device that is implanted under the skin in the chest, back, or abdomen. The pacemaker device can be programmed telemetrically to deliver desired stimulation by

BOX 30.5 (*cont'd*)

adjusting voltage, pulse width, frequency, and electrode configuration.

DBS was used initially for treating tremor associated with Essential Tremor or Parkinson disease, with electrodes placed in the cerebellar-receiving thalamic nucleus (ventral intermediate nucleus; VIM). By the end of the 1990s, DBS was used increasingly for treating other symptoms of Parkinson disease including bradykinesia and rigidity (targets: STN or GPi) and for treating dystonia (target: GPi). DBS can be dramatically effective for these movement disorders.

The mechanism by which DBS works is a subject of investigation. Although the effects are similar to the effect of ablation of the brain target in which the electrodes are placed, DBS does not appear to work simply by inactivating the target nucleus. In fact, there is evidence that DBS activates the efferent axons from that target. Regardless of the mechanism, there is no controversy regarding the efficacy of DBS for specific disorders. Investigations are under way to determine whether other neuropsychiatric disorders such as depression, OCD, and Tourette Syndrome, can be treated safely and effectively with DBS.

Jonathan W. Mink

Suggested Readings

Perlmutter, J. S., & Mink, J. W. (2006). Deep brain stimulation. *Annual Review of Neuroscience, 29,* 229–257.

Wichmann, T., & Delong, M. R. (2011). Deep-brain stimulation for basal ganglia disorders. *Basal Ganglia, 1,* 65–77.

BOX 30.6

OBSESSIVE-COMPULSIVE DISORDER

Obsessive-compulsive disorder (OCD) is a chronic disorder characterized by recurrent intrusive thoughts and ritualistic behaviors that consume much of the afflicted individual's attentional and goal-directed processes. Among the more agonizing illnesses in clinical medicine, OCD is classified as an anxiety disorder because of the marked tension and distress produced by resisting the obsessions and compulsions. Common obsessions involve thoughts of harming oneself or others, of being afflicted by various illnesses, or of being contaminated by germs. Compulsions may be a response to obsessive thoughts (e.g., repetitive hand washing following skin contact with any object due to obsessive worries of having contacted germs) or may instead reflect cognitive-behavioral rituals performed according to stereotyped rules (e.g., counting one's footsteps to avoid ending on certain numbers). Although individuals with OCD recognize that such thoughts and behaviors are irrational, they feel irresistibly compelled to engage in them.

The onset of OCD usually occurs between late childhood and early adulthood. Without treatment, OCD is often disabling. However, chronic treatment with drugs that potently inhibit serotonin reuptake can reduce the amount of time engaged in obsessions and compulsions and the magnitude of the associated anxiety. In addition, behavioral therapy involving repeated *in vivo* exposure and response prevention (e.g., having patients touch a toilet seat and subsequently preventing them from hand washing) may facilitate extinction of the anxiety responses generated by resisting obsessive thoughts and compulsive behaviors.

The etiology and pathophysiology of OCD are unknown. Converging evidence from analysis of the lesions that result in obsessive-compulsive symptoms, functional neuroimaging studies of OCD, and observations regarding the neurosurgical interventions that can ameliorate OCD implicate prefrontal cortical-striatal circuits in the pathogenesis of obsessions and compulsions. The neurological conditions associated with the development of secondary obsessions and compulsions include lesions of the globus pallidus and putamen, Sydenham chorea (a poststreptococcal autoimmune disorder associated with neuronal atrophy in the caudate and putamen), Tourette disorder (an idiopathic syndrome characterized by motoric and phonic tics that may have a genetic relationship with OCD, see Box 30.7), chronic motor tic disorder, and lesions of the ventromedial prefrontal cortex. Several of these conditions are associated with complex motor tics (repetitive, coordinated, involuntary movements occurring in patterned sequences in a spontaneous, unpredictable, and transient manner). Complex tics and obsessive thoughts may thus both reflect aberrant neural processes originating in distinct portions of the cortical-striatal-pallidal-thalamic circuitry that are manifested within the motor and cognitive-behavioral domains, respectively. Compatible with this hypothesis, neurosurgical procedures that interrupt the white matter tracts carrying projections between the frontal lobe, basal ganglia, and thalamus are effective at reducing obsessive-compulsive symptoms in OCD cases that prove intractable to other treatments.

BOX 30.6 (*cont'd*)

Functional neuroimaging studies have also implicated prefrontal cortical-striatal circuits in the pathophysiology of OCD. In primary OCD, "resting" cerebral blood flow and metabolism are increased abnormally in the orbitofrontal cortex and the caudate nucleus, and increase further during symptom provocation in these areas, as well as in the putamen, thalamus, and anterior cingulate cortex. During effective pharmacotherapy, orbital metabolism decreases toward normal, and both drug treatment and behavioral therapy are associated with a reduction of caudate metabolism. In contrast, imaging studies of obsessive-compulsive syndromes arising in the setting of Tourette syndrome or basal ganglia lesions have found reduced metabolism in the orbital cortex in such subjects relative to controls. Differences in the functional anatomical correlates of primary versus secondary OCD suggest that dysfunction arising at multiple points within the ventral prefrontal cortical-striatal-pallidal-thalamic circuitry may result in pathological obsessions and compulsions.

This circuitry in general appears to be involved in the organization of internally guided behavior toward a reward, switching of response strategies, habit formation, and stereotypic behavior. Electrophysiological and lesion analysis studies indicate that parts of the orbitofrontal cortex are specifically involved in the correction of behavioral responses that become inappropriate as reinforcement contingencies change. Such functions could potentially become disturbed at the level of this cortex itself or within the circuits conveying projections from the ventral PFC through the basal ganglia, conceivably giving rise to the perseverative patterns of nonreinforced thought and behavior that characterize the cognitive-behavioral state of OCD.

Wayne C. Drevets

Suggested Readings

Baxter, L. R. (1995). Neuroimaging studies of human anxiety disorders. In F. E. Bloom & D. J. Kupfer (Eds.), *Psychopharmacology: The fourth generation of progress* (Chap. 80, pp. 921–932). New York, NY: Raven Press.

LaPlane, D., Levasseur, M., Pillon, B., Dubois, B., Baulac, M., Mazoyer, B., Baron, J. C., et al. (1989). Obsessive-compulsive and other behavioural changes with bilateral basal ganglia lesions. *Brain, 112,* 69–725.

McDougle, C. J. (1999). The neurobiology and treatment of obsessive-compulsive disorder. In D. S. Charney, E. J. Nestler & B. J. Bunney (Eds.), *Neurobiology of mental illness* (pp. 518–533). New York, NY: Oxford Univ. Press.

BOX 30.7

TOURETTE SYNDROME

A "tic" is a sudden, rapid, recurrent, nonrhythmic, stereotyped movement or sound. Individuals with Tourette syndrome (TS) exhibit prominent and persistent motor tics and at least one phonic tic, beginning in childhood. Current estimates of TS prevalence range between 0.1 and 1.0%, with a 3:1 male:female ratio. Symptoms wax and wane; roughly half of all TS patients experience a significant symptom decline in their early 20s, whereas others experience lifelong symptoms.

Tics differ in complexity (simple, complex) and domain of expression (motor, phonic). In addition to their outward manifestations, a variety of sensory and mental states are associated with tics. "Simple" sensory tics are rapid, recurrent, and stereotyped, and are sensations at or near the skin. "Premonitory urges" are more complex phenomena, which often include both sensory and psychic discomfort that may be momentarily relieved by a tic. The full elaboration of tics, therefore, can include a sequence: (1) a sensory event or premonitory urge, (2) a complex state of inner conflict over if and when to yield to the urge, (3) the motor or phonic production, and (4) a transient sensation of relief.

Functional impairment in TS often results from comorbid conditions. More than 40% of individuals with TS experience symptoms of obsessive-compulsive disorder (OCD), and a comparable number have symptoms of attentional deficits and hyperactivity (ADHD).

Limited neuropathological evidence suggests four locations of potential pathology within cortico-striato-pallido-thalamic (CSPT) circuitry in TS: (1) intrinsic striatal neurons (increased neuronal packing density); (2) diminished striato-pallidal "direct" output (reduced dynorphin-like immunoreactivity in the lenticular nuclei); (3) increased dopaminergic innervation of the striatum (increased density of dopamine transporter sites); and (4) reduced glutamatergic output from the STN (reduced lenticular glutamate content).

BOX 30.7 (*cont'd*)

Volumetric neuroimaging studies report an enlarged corpus callosum, reduced caudate volume, or diminished striatal and lenticular asymmetry; these changes are subtle and are not replicated uniformly. Metabolic neuroimaging studies in TS report a reduced glucose uptake in the orbitofrontal cortex, caudate, parahippocampus, and midbrain regions and reduced blood flow in the caudate nucleus, anterior cingulate cortex, and temporal lobes. Regional glucose uptake patterns suggest distributed CSPT dysfunction, based on covariate relationships between reduced glucose uptake in striatal, pallidal, thalamic, and hippocampal regions. Tic suppression is associated with increased right caudate activation, measured by functional MRI, and by bilaterally diminished activity in the putamen, globus pallidus, and thalamus. Neurochemical imaging studies have reported relatively subtle abnormalities in the levels of dopamine receptors, dopamine release, DOPA decarboxylase, and dopamine transporter in the striatum of some individuals with TS. A report of greater caudate D2 receptor binding among more symptomatic TS identical twins may be relevant to factors contributing to heterogeneity of the TS phenotype.

There are a number of reported subtle abnormalities in the levels of major neurotransmitters, precursors, metabolites, biogenic amines, and hormones in blood, cerebrospinal fluid (CSF), and urine of TS subjects compared to controls. One qualitatively different finding is that of approximately 40% elevations of serum antiputamen antibodies in TS children. This finding may have particular importance, based on growing evidence for autoimmune contributions to at least some forms of TS.

Converging evidence for CSPT pathology in TS comes from neuropsychological and psychophysiological studies. Some forms of TS appear to be accompanied by abnormalities in sensorimotor gating, oculomotor functions, and visuospatial priming consistent with mild corticostriatal dysfunction.

Treatments for TS focus on education, comorbid conditions, and direct tic suppression. Education is a critically important intervention for both family members and affected individuals. Treatments for comorbid conditions, particularly OCD and ADHD, are highly effective and can provide significant relief. Dopamine antagonists, particularly high potency, D2 preferential blockers, are the most potent and rapid acting tic-suppressing agents, but have important, undesirable side effects. Newer, "atypical" antipsychotics are better tolerated, and some are effective in suppressing tics. The DA depleter tetrabenazine also offers significant tic suppression. α_2-Adrenergic agonists are often used as anti-tic agents; compared to dopamine antagonists, these drugs have relatively weaker anti-tic abilities and their benefit generally evolves more gradually. New therapeutic avenues for TS are being explored in controlled studies, including dopamine D1 antagonists, Δ^9-tetrahydro-cannabinol, and deep brain stimulation. Intractable, localized tics have also been treated successfully with injections of botulinum toxic. An effective nonpharmacologic therapy, habit reversal therapy, involves the application of cognitive and behavioral therapy principles to TS, analogous to the successful use of these therapies in the treatment of OCD. Behavioral therapy for tics is as effective as most currently employed pharmacologic therapies.

Neal R. Swerdlow and Jonathan W. Mink

Suggested Readings

Albin, R. L., & Mink, J. W. (2006). Recent advances in Tourette syndrome research. *Trends in Neuroscience, 29,* 175–182.

McNaught, K. S., & Mink, J. W. (2011). Advances in understanding and treatment of Tourette Syndrome. *Nature Reviews Neurology, 8,* 667–676.

Piacentini, J., Woods, D. W., Scahill, L., Wilhelm, S., Peterson, A. L., Chang, S., Walkup, J. T., et al. (2010). Behavior therapy for children with Tourette disorder: A randomized controlled trial. *JAMA, 303,* 1929–1937.

BOX 30.8

THE ROLE OF THE SUBTHALAMIC NUCLEUS IN STOPPING

You are sitting astride your bicycle at an intersection and just about to press down on the pedal when all of a sudden a motorist runs the light. This requires the rapid cancellation of an initiated action. Recent studies suggest that rapid stopping of this kind is implemented by a "hyperdirect" pathway between the frontal cortex and the subthalamic nucleus.

The broader sequence of events that engages this pathway is as follows. Sensory information about the stop signal (in this case, the car) is quickly relayed to the prefrontal cortex, where the stopping command must be generated. Two broad regions of the prefrontal cortex are apparently critical for stopping behavior: the right inferior frontal cortex and the dorsomedial frontal cortex

BOX 30.8 (*cont'd*)

(especially the presupplementary motor area). These two regions appear to work together to send a stop command to intercept the Go process, via the subthalamic nucleus. To appreciate how, first consider the Go process (the initiated pedal press). This is likely generated by premotor areas that project via the direct pathway of the basal ganglia (through striatum, pallidum, and thalamus), eventually exciting primary motor cortex and generating corticospinal volleys to the foot. An important point of interaction for Stop and Go processes could be the globus pallidus. Whereas the Go process activates the striatum and inhibits part of the globus pallidus pars interna, the Stop process activates the subthalamic nucleus which then excites the globus pallidus pars interna, which reduces thalamocortical excitability, leading to increased GABAergic inhibition in primary motor cortex. The consequence is a cessation of the corticospinal volleys for the Go process.

Interestingly, recent studies show that rapidly stopping the hand leads to suppression of foot motor representations and rapidly stopping the voice leads to suppression of hand motor representations. This shows that stopping has very broad effects on the motor system (a "global stop command"). While not proven, this could relate to the fact that the subthalamic nucleus appears to have a very broad effect on multiple neurons in the globus pallidus pars interna (Fig. 30.8).

Although rapid stopping has only so far been explored in the action domain, note that the subthalamic nucleus has different territories, including sensorimotor, associative and limbic, which connect to motor/premotor cortical, lateral prefrontal and cingulate/orbital frontal regions, respectively. It is an open question whether the kind of stopping discussed here could extend, for example, to the limbic domain of the subthalamic nucleus to enable emotional and motivational control. Notably, the limbic domain of the subthalamic nucleus is now a target for deep brain stimulation in obsessive compulsive disorder, and some studies have reported that inadvertent stimulation of this area in Parkinson's patients induces hypomania.

Other research points to the importance of neural oscillations in the beta frequency band for the stopping system. Successfully stopping action is associated with increased beta power in the subthalamic nucleus, as well as the presupplementary motor area and the right inferior frontal cortex. It is possible that stopping is implemented via long-distance synchronization in the beta band in the overall cortico-basal ganglia network.

Adam R. Aron

Suggested Readings

Aron, A. R., Durston, S., Eagle, D. M., Logan, G. D., Stinear, C. M., & Stuphorn, V. (2007). Converging evidence for a fronto-basal-ganglia network for inhibitory control of action and cognition. *Journal of Neuroscience, 27,* 11860–11864.

Chambers, C. D., Garavan, H., & Bellgrove, M. A. (2009). Insights into the neural basis of response inhibition from cognitive and clinical neuroscience. *Neuroscience and Biobehavioral Reviews, 33,* 631–646.

Ray, N., Brittain, J., Holland, P., Joundi, R., & Stein, J. (2012, March). The role of the subthalamic nucleus in response inhibition: Evidence from local field potential recordings in the human subthalamic nucleus. *Neuroimage, 60*(1), 271–278.

Summary

It is appropriate to consider the basal ganglia as part of the motor system because the largest portion of the basal ganglia is devoted to motor control and the most prominent deficits resulting from basal ganglia damage are motor. However, it has become clear that the basal ganglia play substantial roles in nonmotor function and in more cognitive aspects of movement. These roles are the subject of active research. Ultimately there may be a unifying theory of basal ganglia function that will encompass all of the subdivisions, but at this point one can only say that it is likely that the different circuits employ similar mechanisms.

References

Albin, R. L., Young, A. B., & Penney, J. B. (1989). The functional anatomy of basal ganglia disorders. *Trends in Neurosciences, 12*(10), 366–375.

Alexander, G. E., & Crutcher, M. D. (1990). Functional architecture of basal ganglia circuits: Neural substrates of parallel processing. *Trends in Neurosciences, 13*(7), 266–271.

Aosaki, T., Tsubokawa, H., Ishida, A., Watanabe, K., Graybiel, A. M., & Kimura, M. (1994). Responses of tonically active neurons in the primate's striatum undergo systematic changes during behavioral sensorimotor conditioning. *Journal of Neuroscience, 14*(6), 3969–3984.

Benecke, R., Rothwell, J. C., Dick, J. P., Day, B. L., & Marsden, C. D. (1986). Performance of simultaneous movements in patients with Parkinson's disease. *Brain, 109,* 739–757.

Flaherty, A. W., & Graybiel, A. M. (1991). Corticostriatal transformations in the primate somatosensory system: Projections from phsyiologically mapped body-part representations. *Journal of Neurophysiology, 66*(4), 1249–1263.

Gage, G. J., Stoetzner, C. R., Wiltschko, A. B., & Berke, J. D. (2010). Selective activation of striatal fast-spiking interneurons during choice execution. *Neuron, 67,* 466–479.

Gittis, A. H., Leventhal, D. K., Fensterheim, B. A., Pettibone, J. R., Berke, J. D., & Kreitzer, A. C. (2011). Selective inhibition of striatal fast-spiking interneurons causes dyskinesias. *Journal of Neuroscience, 31,* 15727–15731.

Haber, S. N., Fudge, J. L., & McFarland, N. R. (2000). Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorso-lateral striatum. *Journal of Neuroscience, 20,* 2369–2382.

Hikosaka, O., Takikawa, Y., & Kawagoe, R. (2000). Role of the basal ganglia in the control of purposive saccadic eye movements. *Physiological Reviews, 80,* 953–978.

Hikosaka, O., & Wurtz, R. H. (1985). Modification of saccadic eye movements by GABA-related substances. II. Effects of muscimol in monkey substantia nigra pars reticulata. *Journal of Neurophysiology, 53,* 292–308.

Hoover, J. E., & Strick, P. L. (1993). Multiple output channels in the basal ganglia. *Science, 259,* 819–821.

Jog, M., Kubota, Y., Connally, C. I., Hillegart, V., & Graybiel, A. M. (1999). Building neural representations of habits. *Science, 286,* 1745–1749.

Kelly, R., & Strick, P. L. (2004). Macro-architecture of basal ganglia loops with the cerebral cortex: Use of rabies virus to reveal multisynaptic circuits. *Progress in Brain Research, 143,* 449–459.

Marsden, C. D. (1987). What do the basal ganglia tell premotor cortical areas? *Ciba Foundation Symposium, 132,* 282–300.

Mink, J. W. (2003). The basal ganglia and involuntary movements: Impaired inhibition of competing motor patterns. *Archives of Neurology, 60,* 1365–1368.

Mink, J. W., & Thach, W. T. (1991). Basal ganglia motor control. III. Pallidal ablation: normal reaction time, muscle cocontraction, and slow movement. *Journal of Neurophysiology, 65,* 330–351.

Parent, A., & Hazrati, L. -N. (1993). Anatomical aspects of information processing in primate basal ganglia. *Trends in Neurosciences, 16*(3), 111–116.

Redgrave, P., Prescott, T., & Gurney, K. (1999). The basal ganglia: A vertebrate solution to the selection problem? *Neuroscience, 89,* 1009–1023.

Romo, R., Scarnati, E., & Schultz, W. (1992). Role of primate basal ganglia and frontal cortex in the internal generation of movements. II: Movement-related activity in the anterior striatum. *Experimental Brain Research, 91,* 385–395.

Tremblay, L., Hollerman, J., & Schultz, W. (1998). Modifications of reward expectation-related neuronal activity during learning in primate striatum. *Journal of Neurophysiology, 80,* 964–977.

Wilson, C. J., & Groves, P. M. (1980). Fine structure and synaptic connections of the common spiny neuron of the rat neostriatum: A study employing intracellular injection of horseradish peroxidase. *The Journal of Comparative Neurology, 194,* 599–614.

Suggested Readings

Bolam, J. P., Hanley, J. J., Booth, P. A., & Bevan, M. D. (2000). Synaptic organisation of the basal ganglia. *Journal of Anatomy, 196,* 527–542.

Graybiel, A. (1998). The basal ganglia and chunking of action repertoires. *Neurobiology of Learning and Memory, 70,* 119–136.

Houk, J. C., Davis, J. L., & Beiser, D. G. (Eds.)., *Models of information processing in the basal ganglia: Computational neuroscience*. Cambridge, MA: MIT Press.

Kreitzer, A. C. (2009). Physiology and pharmacology of striatal neurons. *Annual Review of Neuroscience, 32,* 127–147.

Middleton, F., & Strick, P. (2000). Basal ganglia output and cognition: Evidence from anatomical, behavioral, and clinical studies. *Brain and Cognition, 42,* 183–200.

Mink, J. W. (1996). The basal ganglia: Focused selection and inhibition of competing motor programs. *Progress in Neurobiology, 50,* 381–425.

Penney, J. B., & Young, A. B. (1983). Speculations on the functional anatomy of basal ganglia disorders. *Annual Review of Neuroscience, 6,* 73–94.

Watts, R. L., & Roller, W. C. (1997). *Movement disorders: Neurologic principles and practice*. New York, NY: McGraw-Hill.

Jonathan W. Mink

CHAPTER

31

Cerebellum

The cerebellum is a softball-sized structure located at the base of the skull. Grab the back of your head just above where it meets your neck: your hand is now cupped around your cerebellum. As with most brain systems, much of what we know about the cerebellum stems from symptoms of damage or pathology and from its connectivity with the rest of the brain. From such evidence it has long been clear that the cerebellum is an important component of the motor system. Severe abnormalities of movement are produced by pathologies of the cerebellum. Cerebellar output influences systems that are unambiguously motor, and inputs to the cerebellum convey information known to be essential for movement such as joint angles and loads on muscles. Recently it has become equally clear that the role of the cerebellum is not limited to movement. This is indicated by its interconnections with nonmotor structures; by the more subtle, nonmotor deficits seen with cerebellar lesions; and by functional imaging studies where regions of the cerebellum show activation during nonmotor tasks. A deeper understanding of the cerebellum will require identification of the common aspects or computational demands that these motor and nonmotor functions share.

Research on the anatomy, physiology, and function of the cerebellum is complemented by computational analyses of the rules for input/output transformations and the means by which the cerebellar circuit implements such transformations. This interdisciplinary approach was made possible by the seminal work of Eccles, Ito, and Szentágothai, who in 1967 published *The Cerebellum as a Neuronal Machine*, which described most of the essential aspects of the cellular and synaptic organization of the cerebellum. Two years later, David Marr published a remarkable paper in which he inferred some basic computational properties of the cerebellum solely on the basis of the wiring diagram that Eccles and associates had elucidated. The relative clarity of the cerebellar circuit and the close interconnections between cerebellar activity and motor output have enabled rigorous experimental and computational analyses of the relationship between brain and behavior, and even between circuit plasticity and learning.

CEREBELLAR ANATOMY AND CIRCUIT

The anatomical, circuit, and physiological characteristics of the cerebellum suggest that it carries out a different type of computation than the cerebral cortex. One indication of this distinction is that Purkinje cells, which provide the only output of the cerebellar cortex, are inhibitory (GABAergic) and fire about 50 action potentials every second, even in the absence of sensory input. Their downstream target neurons in the vestibular and cerebellar nuclei, which comprise the sole output of the entire cerebellum, are also spontaneously active. High tonic firing rates enable these cerebellar cell types to respond rapidly, precisely, and bidirectionally to input signals. As a result, the cerebellum can exert tight temporal control over its diverse downstream target regions. This chapter first discusses cerebellar anatomy, circuit, and development, and then examines the cellular and synaptic specializations that underlie cerebellar function. The last section demonstrates the relationship between circuit and function during behavior and learning.

Phylogenetic Development

The cerebellum is present in all vertebrates. In agnathans (lampreys and hagfish), it is a rudimentary structure that assists the functions of the well-developed vestibulo-ocular, vestibulospinal, and reticulospinal systems. The cerebellum is somewhat larger in fishes, where, on the input side, it processes sensory information from the vestibular,

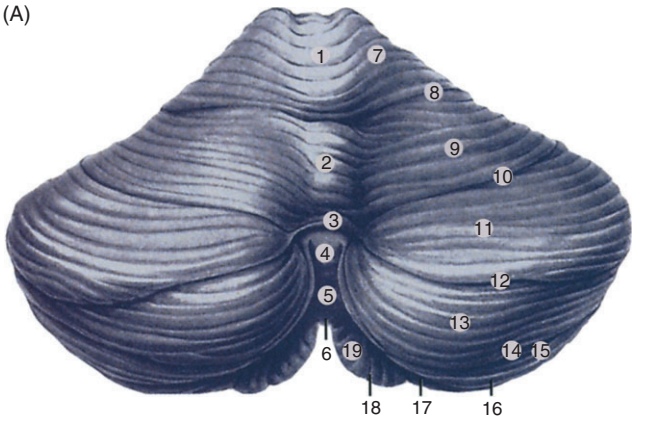
© 2013 Elsevier Inc. All rights reserved.

lateral line, and to a lesser extent proprioceptive and somatosensory systems. On the output side it is connected to the vestibular and reticular nuclei. In amphibians the region of the cerebellum that receives proprioceptive and other sensory information is expanded, and it increases still further in reptiles, birds, and most mammals, where it constitutes the largest portion of the cerebellum. A diverse range of brain regions send inputs to or receive outputs from the cerebellum. In primates and dolphins, the hugely expanded lateral hemispheres are connected with the enlarged cerebral cortex via the thalamus and pons.

Gross Anatomy

The cerebellum consists of a three-layered cortex folded repeatedly, accordion-style, around the three pairs of cerebellar nuclei buried in the middle. Each fold of the cerebellar cortex is called a folium (leaf), and in most species the folia run roughly transverse to the long axis of the body, from ear to ear. The folia are grouped into lobules, numbered I to X; lobulation is fairly consistent across individuals of the same species and even across species, despite great variation in cerebellar development. The lobules in turn are grouped into three lobes. The flocculonodular lobe (also known as the vestibulocerebellum) is the phylogenetically oldest portion and is located on the inferior surface. Above the posterolateral fissure sit the anterior and posterior lobes, separated from each other by the primary fissure (Fig. 31.1). If the entire human cerebellar cortex were unfolded and flattened, it would have about the same area as a sheet of paper, but arranged in a long thin strip.

The cerebellar cortex influences the rest of the brain exclusively via projections to neurons in the vestibular and cerebellar nuclei (also termed the "deep cerebellar nuclei"). From medial to lateral, the cerebellar nuclei comprise the fastigius, the interpositus (further divided into the globose and emboliform in humans), and the dentate. The cerebellar cortex, excluding the flocculonodular lobe, can be divided into three longitudinal (parasagittal) zones based on the projection of the cortex onto the three cerebellar nuclei (Fig. 31.2). The medial zone receives information from a wide range of sensory systems, including vestibular, visual, somatosensory, and auditory. The medial zone also processes signals relating to autonomic, visceral, cardiovascular, and immune function. It projects to the fastigial nuclei, which in turn send outputs to the vestibular and reticulospinal systems, as well as the thalamus and superior colliculus. The intermediate zone of the cerebellar cortex receives proprioceptive and somatosensory information from the spinal cord, as well as information from the motor cortex via the pontocerebellar nuclei. Its axons target the interpositus nuclei, which in turn have a major



1. Culmen
2. Declive
3. Vermal folium
4. Vermal tuber
5. Vermal pyramis
6. Vermal uvula
7. Quadrangular lobule
8. Primary fissure
9. Simplex lobule
10. Superior posterior fissure
11. Superior semilunar lobule
12. Horizontal fissure
13. Inferior semilunar lobule
14. Ansoparamedian fissure
15. Gracile lobule
16. Prebiventral fissure
17. Biventral lobule
18. Secondary fissure
19. Cerebellar tonsil

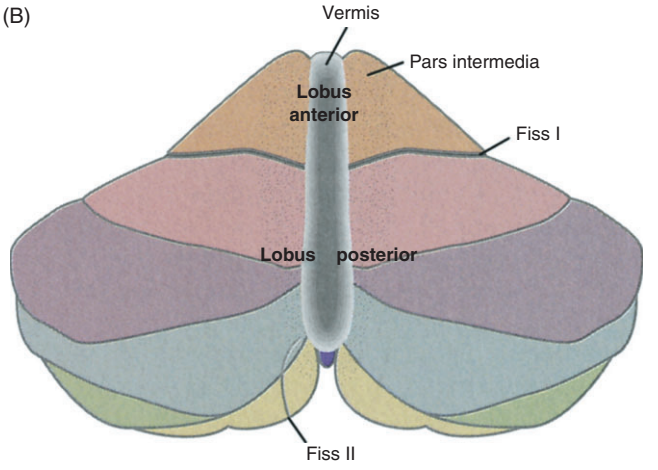


FIGURE 31.1 Cerebellar structure and functional subdivisions. (A) Dorsal view of the human cerebellum. (B) Functional subdivisions. *Adapted from Nieuwenhuys, Voogd, and van Huijzen (1979).*

projection to the red nucleus, with some axons continuing to the contralateral thalamus (VL). There are also a sparser number of collaterals projecting to the spinal cord and pons. The lateral zone receives diverse types of information from the pontine nuclei, which carry signals from the widespread regions of the cerebral cortex. The lateral zone projects to the dentate nuclei, whose predominant output is to the thalamus (VL) and reticular formation, with lesser outputs to the red nucleus and spinal cord (Fig. 31.2). Finally, the flocculonodular lobe receives vestibular, visual, and motor information, and sends its output exclusively to the brainstem vestibular nuclei. In addition to the premotor outputs shown in Figure 31.2, each ouput nucleus also sends inhibitory outputs to subregions of the inferior olive: vestibular and fastigial nuclei to

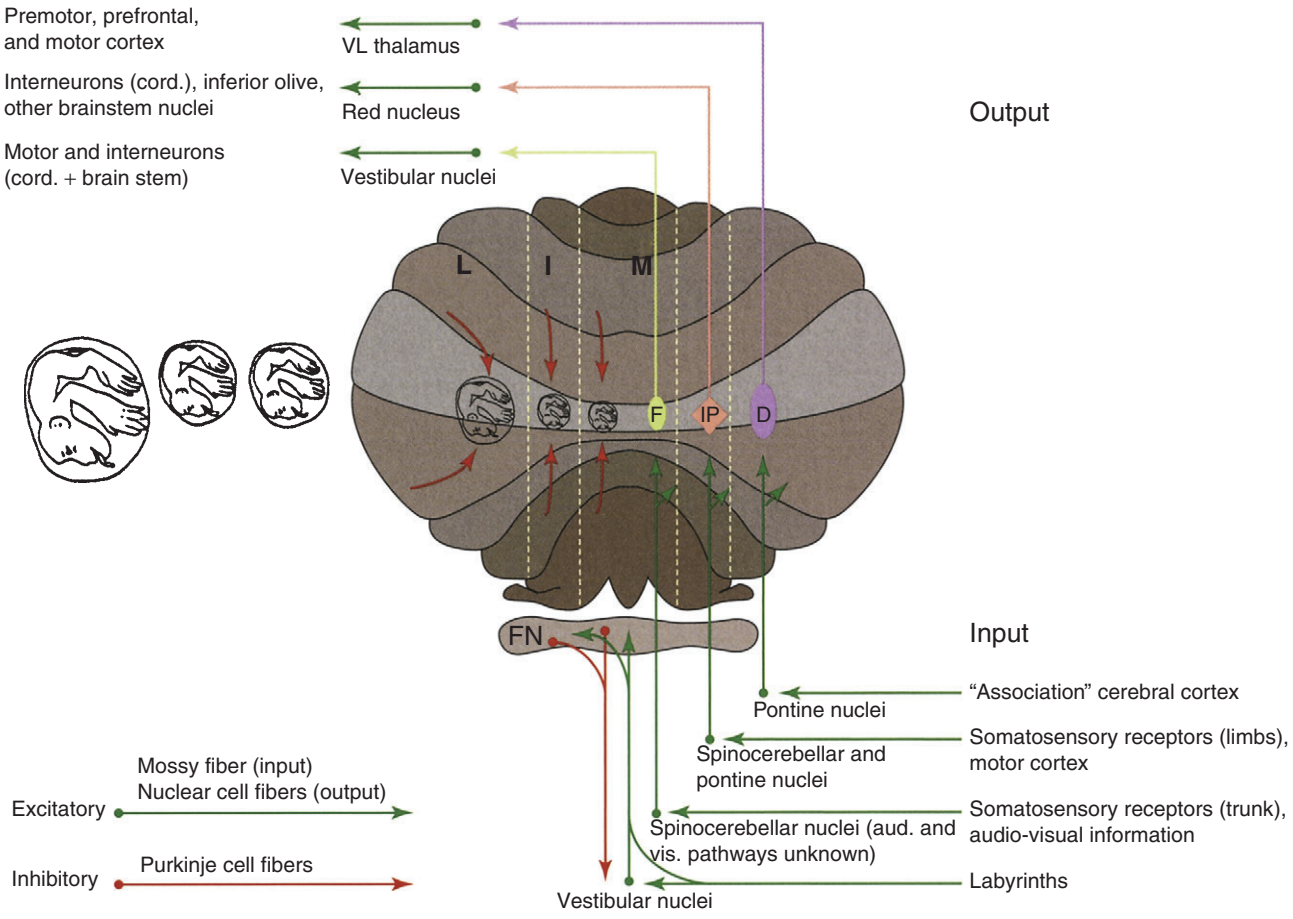


FIGURE 31.2 Cerebellar inputs and outputs. The cerebellum is shown schematically with the flocculonodular (FN) lobe unfolded and represented at the inferior surface. Green arrows indicate excitatory projections into and out of the cerebellum; red arrows indicate the inhibitory projections from cerebellar cortex to the deep cerebellar nuclei and the vestibular nuclei. *Adapted from Thach, Fig. 31-3. In Mountcastle, J. B. (1980).* Medical Physiology, *Volume I. 14 Ed. C. V. Mosby Co., St. Louis.*

the dorsal cap of Kooy, the beta nucleus, and weakly to the medial accessory olive; interpositus to the dorsal and medial accessory olive, with a weaker projection to the cap of Kooy; and the dentate to the principal olive.

Mapping studies reveal that the cerebellum is not organized purely somatotopically like the sensory cerebral cortices. Part of the difficulty in defining cerebellar organizing principles is that the cerebellum combines diverse sensory inputs to create and refine motor commands. Thus, neither sensory nor motor mapping individually may square exactly with cerebellar organization. Sensory mapping studies reveal a "fractured somatotopy," in which the regions of cerebellum that respond to touch of paw sites, for example, are located close to each other, but sensation for forelimb is represented elsewhere. The cerebellum can also be partitioned into parasagittal zones that may represent distinct motor control modules (see Box 31.1).

Two different schema have been proposed to account for cerebellar organization. In the first, each major subdivision of the cerebellum controls all parts of the body, but during different behavioral requirements: the flocculonodular lobe and medial cerebellum modulate vestibular-driven behaviors and eye movements; the intermediate zone controls somatic reflexes; and the lateral zone concerns itself with movements under voluntary control. In a second schema, the medial zone primarily controls musculature of the trunk, the intermediate zone that of the proximal limbs, and the lateral zone the distal limbs. Both of these organizing principles are supported by experimental and clinical studies in which the different cerebellar nuclei are lesioned or inactivated (Fig. 31.3). Neither accounts for the role of the cerebellum in regulating autonomic, emotional, or cognitive functions. The inputs and outputs of the cerebellum are diverse, and the outputs are relatively poorly characterized to date. A unifying view of cerebellar function that draws on analyses of neuronal activity, circuits, and behavior remains to be established.

Basic Cerebellar Circuit

The structure of the cerebellar circuit is highly stereotyped (Fig. 31.4; though see Box 31.2). Mossy fiber inputs,

BOX 31.1

PARASAGITTAL STRIPES: THE BASIS OF CEREBELLAR MICROCIRCUITS?

One of the greatest unknowns about the cerebellar circuit is whether the prototypical olivocerebellar loop, from olive → Purkinje cells → cerebellar nuclei → back to olive, is actually organized so that an *individual* olivary neuron receives inhibitory feedback from the same inhibitory nucleus neurons that it influences indirectly via Purkinje cells. In other words, does the microcircuit (at the level of individual neurons) reflect the macrocircuit (at the level of entire structures)? The most suggestive work on this question comes from anatomical studies on the parasagittal zones of the cerebellum. Histological experiments have revealed that many proteins, including aldolase C, acetylcholinesterase, calbindin, and parvalbumin, are expressed differentially in parasagittal stripes that run the length of each longitudinal zone. These stripes, which have been given names (1+, 1−, 2+, 2−, etc.), are constant in their presence and location within a species and to some extent across species (Pakan, Graham, Gutiérrez-Ibáñez, & Wylie, 2011).

Recent work has highlighted the potential significance of these stripes in organization of microcircuits in the cerebellum. Each olivary neuron gives rise to 7 to 10 climbing fibers in the cortex; all of the Purkinje cells that receive input from one olivary neuron are located within a stripe defined by aldolase C staining. Even in cases where the olivary axon diverges to different cerebellar lobules, the climbing fibers in both lobules obey the pattern established by aldolase C stripes: either all branches are in positive stripes or all in negative stripes. Furthermore, a group of Purkinje cells that are innervated by a small cluster of olivary neurons typically exhibit converging axon arbors in the cerebellar nuclei as well, within a subregion that is also defined by aldolase C staining (Sugihara, 2011). This convergence holds true even when the Purkinje cells are located in separate lobules. The axon collaterals of olivary neurons, which extend into the cerebellar nuclei, also form small clusters of terminals suggestive of microcircuit targeting. In turn, a cluster of nuclear neurons appears to target the same olivary region from which it receives input. Even cortical interneurons may respect the boundaries defined by aldolase stripes (Sillitoe & Joyner, 2007). Together these data are strongly suggestive of a closed loop olivo-cortical-nuclear microcircuit.

These findings are exclusively anatomical, but some experiments have extended these concepts to physiology and function as well. For example, muscle injection of a rabies virus tracer, which can travel retrogradely across many synapses, reveals one or more longitudinal stripes of Purkinje cells bounded by aldolase C compartments. Physiologically, Purkinje cells in aldolase positive and negative stripes express different complements of glutamate transporters, some of which are more effective than others. As a result, it is easier to trigger metabotropic glutamate receptor activity, and the consequent cascade leading to long-term depression of the parallel fiber → Purkinje cell synapse, in aldolase-negative than in aldolase-positive compartments (Wadiche & Jahr, 2001). In sum, the parasagittally organized stripes of the cerebellum are highly significant for both circuit and function.

Martha Bagnall

References

Pakan, J. M., Graham, D. J., Gutiérrez-Ibáñez, C., & Wylie, D. R. (2011). Organization of the cerebellum: correlating zebrin immunohistochemistry with optic flow zones in the pigeon flocculus. *Visual Neuroscience, 28,* 163–173.

Sillitoe, R. V., & Joyner, A. L. (2007). Morphology, molecular codes, and circuitry produce the three-dimensional complexity of the cerebellum. *Annual Review of Cell and Developmental Biology, 23,* 549–577.

Sugihara, I. (2011). Compartmentalization of the deep cerebellar nuclei based on afferent projections and aldolase C expression. *Cerebellum, 10,* 449–463.

Wadiche, J. I., & Jahr, C. E. (2001). Multivesicular release at climbing fiber-Purkinje cell synapses. *Neuron, 32,* 301–313.

which are excitatory, arise from a wide variety of pontine, medullary, and spinal areas (Fig. 31.2). Mossy fibers diverge to send axon collaterals to the cerebellar nuclei before ascending to the inner, granular layer of the cerebellar cortex; they typically synapse bilaterally but in some cases are restricted ipsilaterally. In each folium of the cortex, a mossy fiber makes up to 50 large glomerular synapses with a relatively large presynaptic specialization known as a *rosette*, which excites one dendrite from each of ~20–30 granule cells. Granule cells typically have 3 to 5 short dendrites, each of which terminates in a glomerulus; thus, each granule cell receives input from 3 to 5 mossy fibers. Granule cells are the most common cell type in the entire brain (~50 billion in humans), providing an unparalleled capacity for combinatorial and sparse coding of inputs. The axons of granule cells ascend radially, making a few synaptic contacts on Purkinje cells on their way to the outer molecular layer, where they bifurcate into a T shape and extend long, unmyelinated branches termed the parallel fibers (so named because they run transversely, parallel to the folia). Each parallel fiber reaches up to 1 cm through the cortex and releases glutamate onto the

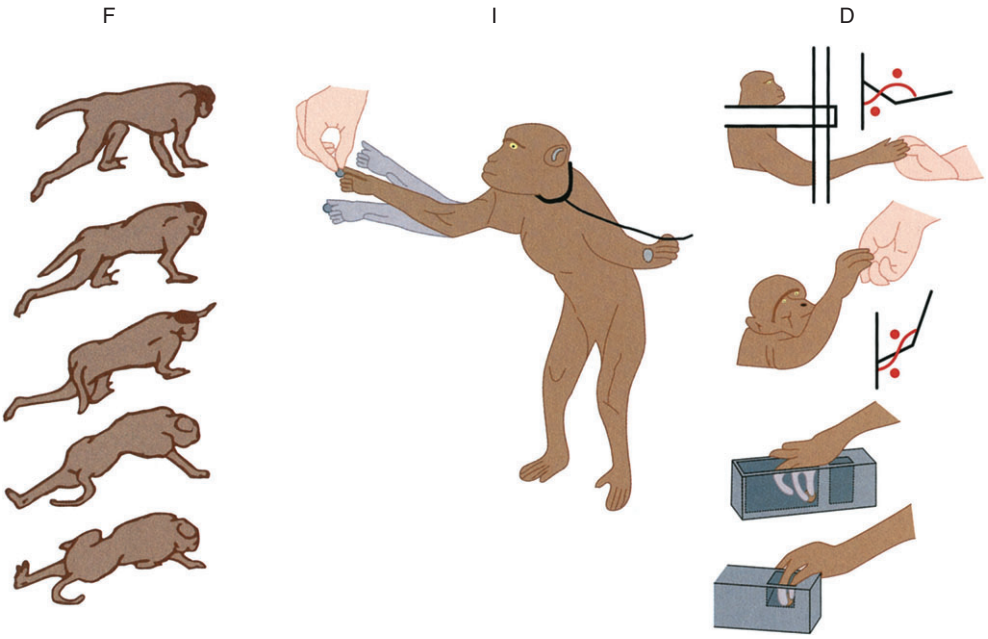


FIGURE 31.3 Major deficits produced by microinjection of muscimol and kainic acid into different regions of the cerebellum. F indicates the standing and walking deficit seen after muscimol injections into the fastigius. I indicates arm position (tremor) during reaching after muscimol injection into the interpositus. D indicates the deficits in reaching and pinching after muscimol injection into the dentate. *Adapted from Thach et al. (1980).*

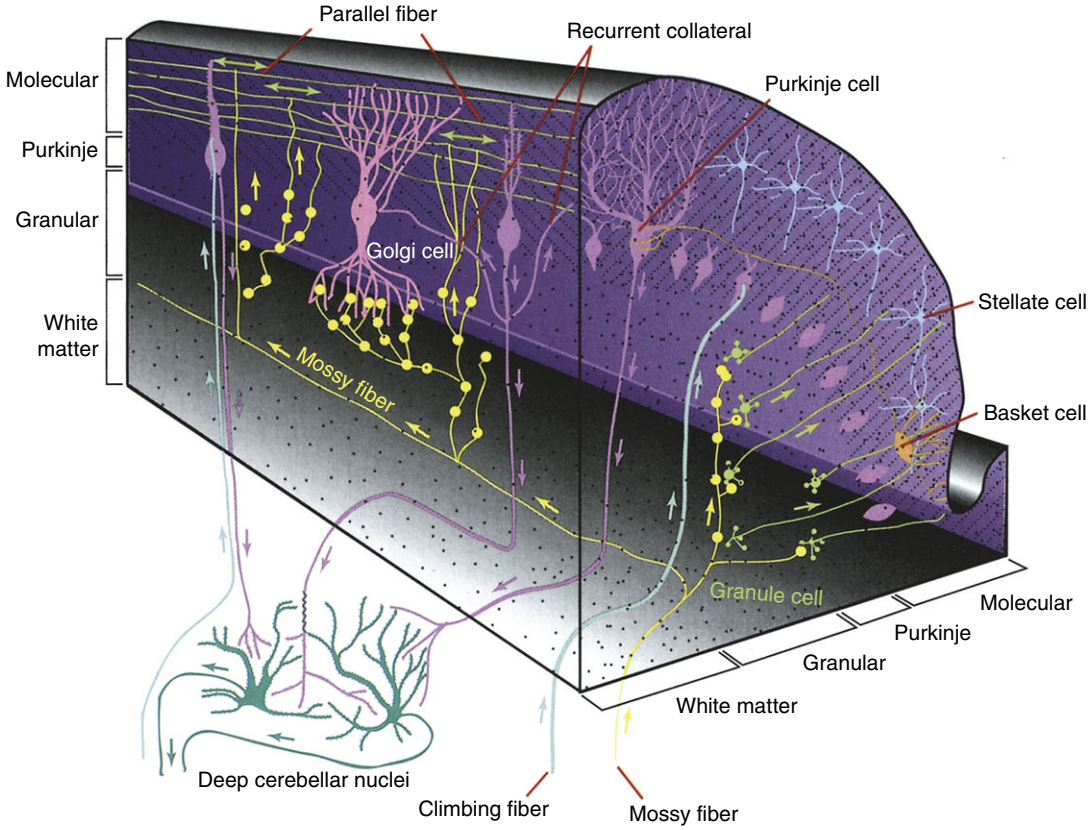


FIGURE 31.4 Schematic view of cerebellar cortex. This section shows the three-dimensional relationship between the different cell types—specifically, the medullary layer or white matter, granular layer, basket cells, stellate cells, Purkinje cells, Golgi cells, granular cells, parallel fibers, climbing fibers, mossy fibers, deep cerebellar nuclei, and recurrent collaterals. *Adapted from Fox, C. A. (1962). The structure of the cerebellar cortex. In* Correlative Anatomy of the Nervous System. *New York, MacMillan. [Fig. 16.10A].*

BOX 31.2

DIFFERENCES BETWEEN THE VESTIBULAR AND NONVESTIBULAR PORTIONS OF THE CEREBELLUM

Phylogenetically, the vestibular and fastigial (medial) cerebellar nuclei predate the interpositus and dentate. Perhaps as a result, the vestibular and fastigial cerebellar circuits exhibit some distinctive properties compared to their relatively younger neighbors:

1. Unipolar brush cells are present in highest densities in the granule cell layer of the flocculonodular lobe. These excitatory neurons receive mossy fiber input, like granule cells, but synapse locally onto granule cells. Golgi cells feed back to unipolar brush cells with mixed glycinergic and GABAergic synapses, in contrast to the purely GABAergic feedback to granule cells (Dugue, Dumoulin, Triller, & Dieudonne, 2005). The role of these neurons in the vestibular circuit is unclear.
2. Olivary neurons in the dorsal cap of Kooy, to which both the vestibular and fastigial nuclei project, have different intrinsic electrical properties than those in the principal and accessory olive. Neurons in the dorsal cap exhibit an approximately linear relationship between injected current and resulting firing rate, tonically firing at up to 35 Hz during a current step. In contrast, neurons in the rest of the olive exhibit significant subthreshold oscillations, usually only fire a couple of spikes at a maximum of 10 Hz in response to depolarization, and are relatively insensitive to depolarizing current injection (Urbano, Simpson, & Llinas, 2006).
3. In a similar vein, the GABAergic feedback from the vestibular cerebellum to the dorsal cap of Kooy is phasic and time-locked to presynaptic firing. The comparable feedback from nonvestibular cerebellum to the principal and accessory olive is slow, aphasic, and appears to be mediated by asynchronous release of GABA (Best & Regehr, 2009).
4. The vestibular and fastigial nuclei contain two major types of premotor projection neurons: glutamatergic neurons that project to contralateral targets, and glycinergic neurons that project to ipsilateral targets. Premotor projection neurons in the interpositus and dentate are exclusively contralateral-projecting and glutamatergic (although the dentate nucleus contains in addition some nucleocortical glycinergic neurons of unknown role) (Bagnall et al., 2009; Shin et al., 2011; Uusisaari & Knopfel, 2010).
5. Some projection neurons in the vestibular nuclei synapse directly onto motor neurons, whereas projection neurons in the nonvestibular cerebellum influence behavior less directly via synapses onto thalamus, pons, medulla, and colliculi.

As a result of these differences, it is not known whether computations carried out by one portion of the cerebellum are necessarily the same as those carried out by the other. The generally homologous structures of the vestibular and nonvestibular parts of the cerebellum may mask significant underlying dissimilarities in the nature of their information processing. This distinction may be related to the need for high-speed processing in the vestibular system, or perhaps to the relatively tight coupling between vestibular signaling and motor output.

Martha Bagnall and Michael Mauk

dendritic spines of ~75 Purkinje cells. The parallel fibers in any one region of the cortex may arise from granule cells in the medial, intermediate, and lateral zones of the cerebellum. Thus, each Purkinje cell may receive information about sensory conditions, internal states, and the plans of the organism. It is notable that while the synapse from mossy fiber to granule cell is robust, each parallel fiber synapse on Purkinje cells is quite small (1 nS measured at soma). Therefore, it is assumed that the convergent activity of many parallel fibers is required to modulate the simple spike firing rate of Purkinje cells. Approximately 150,000 parallel fibers contact a given Purkinje cell.

The second major input to the cerebellum is the climbing fiber, which originates from inferior olivary neurons in the contralateral brainstem. Climbing fibers branch in the sagittal plane and make contact with 7 to 10 Purkinje cells, often in more than one lobule. After a developmental pruning period, each Purkinje cell receives input from only one climbing fiber, which makes about 300 synaptic contacts onto its soma and dendrites. The simultaneous release of multiple vesicles of glutamate from each of these contacts results in a powerful postsynaptic current that drives an unusual action potential termed a "complex spike." The complex spike consists of an action potential followed by several spikelets riding on the back of a larger plateau depolarization (Fig. 31.5). Two to four of these spikelets are sufficiently large to drive axonal action potentials, so that a downstream neuron in the cerebellar or vestibular nuclei is thought to see a short, high-frequency burst of Purkinje input after each complex spike. Olivary neurons fire at about 1 Hz and increase firing as an animal encounters novel conditions or tries to adapt its movement.

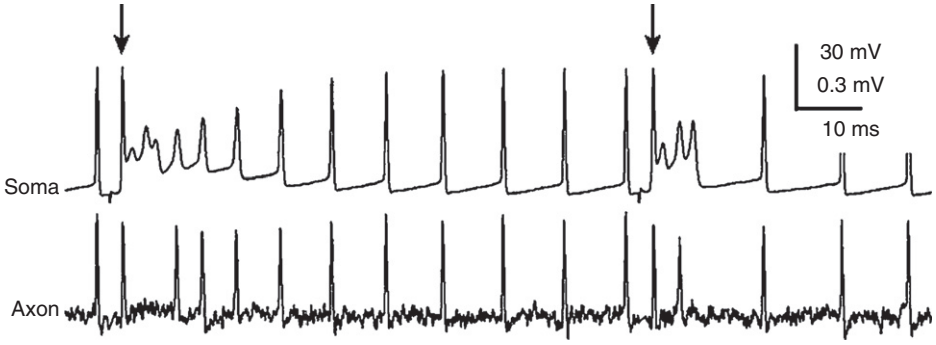


FIGURE 31.5 Whole-cell recording of a Purkinje cell at the soma (top) and cell-attached recording from its axon (bottom). The Purkinje cell fires "simple spikes" spontaneously, all of which are transmitted down the axon. At the times indicated by arrows, a complex spike is elicited by climbing fiber stimulation. Two to three of the spikes associated with the complex spike drive axonal spiking. *Adapted from Khaliq and Raman (2005).*

Most climbing fiber axons also emit collaterals in the cerebellar nuclei, but these synaptic connections have only been studied to date at the anatomical, not physiological, level.

Purkinje cell output is inhibitory and targets exclusively the neurons of the cerebellar and vestibular nuclei. The axon of a Purkinje cell projects ipsilaterally to one of these nuclei, where it diverges onto three main types of neurons: (1) GABAergic neurons that project back to the inferior olive, closing an inhibitory feedback loop; (2) large projection neurons, most of which are excitatory; and (3) short-axon inhibitory neurons of unknown function. The large projection neurons serve as the sole output of the cerebellar circuit and send their axons to regions of the brain as diverse as the thalamus, superior colliculus, and pons. A small subset project back to the cerebellar cortex, with unknown functional consequences (Gould & Graybiel, 1976; Uusisaari & Knopfel, 2010). Surprisingly, given their role as sole output of the large and computationally intense cerebellum, cerebellar and vestibular nucleus neurons are few in number, approximately 1/30 as many as Purkinje cells (Palkovits, Mezey, Hamori, & Szentagothai, 1977). Similar to Purkinje cells, they fire spontaneously at high rates *in vivo* and *in vitro* (~30 Hz, with significant variation), with their firing rates modulated during various types of behaviors.

Complete Cerebellar Circuit

Five other cell types make up the complete cerebellar circuit (Fig. 31.4). Inhibitory Golgi cells reside in the upper portion of the granule cell layer. Their ascending dendrites extend to the outer molecular layer, where they receive excitatory inputs from parallel fibers; their basal dendrites remain in the granule cell layer, where they receive excitatory inputs from mossy fibers. Each Golgi cell synapses onto several thousand granule cells, enabling it to provide both feedforward and feedback inhibition, depending on whether its activity is driven more by mossy fibers or parallel fibers, respectively. Golgi cells also appear to make direct inhibitory synapses on each other (Hull & Regehr, 2012).

Also in the granule cell layer are Lugaro cells, located just below the Purkinje cell layer. Lugaro cells form inhibitory synaptic contacts onto Golgi cells as well as basket and stellate cells in the outer molecular layer (see below). A distinguishing characteristic is that Lugaro cells, unlike Golgi cells, are sensitive to serotonin. Serotonergic activation of Lugaro cells may indirectly activate other circuit elements by inhibiting inhibitory Lugaro cells (Geurts, De Schutter, & Dieudonne, 2003).

The third identified member of the granule cell layer is the glutamatergic unipolar brush cell (Mugnani, Sekerková, & Martina, 2011), which is found in highest densities in vestibular regions (the nodulus and flocculus) of the cerebellum. Like granule cells, they receive excitatory inputs from mossy fibers, but their axons branch locally within the granule cell layer and form excitatory synapses onto granule cells. They are inhibited by Golgi cells at glycinergic or mixed glycinergic/GABAergic synapses. *In vitro* studies have shown that unipolar brush cells respond to brief bursts of excitatory input with prolonged bursts of output that far outlast the stimulus. The influence of this cell type on network computation is far from understood.

Two types of GABAergic interneurons, basket and stellate cells, reside in the outer molecular layer. Both are excited by parallel fibers and inhibit Purkinje cells; basket cells target the soma and perisomatic region of the Purkinje cell, whereas stellate cells contact Purkinje dendrites. Thus, both basket and stellate cells provide feedforward inhibition from granule cells to Purkinje cells. Both cell types also receive weaker, nonsynaptic excitation from climbing fibers. Basket cells, but not stellate cells, are inhibited by Purkinje cell axon collaterals that ramify near the Purkinje cell layer, creating an inhibitory loop. Because the basket cell axons traverse along parasagittal strips, this inhibitory loop may provide a way for each Purkinje cell in a parasagittal strip to entrain the firing of its sister Purkinje cells via disinhibition (inhibition of inhibition). Classically, these outer molecular layer interneurons are thought to provide a type of "off-beam" inhibition, wherein the parallel fibers excite Purkinje cells directly in their path, but Purkinje cells located to either side are inhibited by basket and stellate cells. Thus, the net effect of activity in a small cluster of granule cells would be an increase in firing in the transverse strip of Purkinje cells directly above

and a decrease in firing in the strips of Purkinje cells on either side of the "on-beam" activity. Whether this occurs outside of the context of experiments stimulating beams of parallel fibers and whether basket and stellate cells contribute equally to this phenomenon are both unclear.

Finally, in addition to the mossy fiber and climbing fiber inputs to the cerebellum, there is a third type of modulatory input with a widespread, nonlaminar distribution. These afferents originate in the locus ceruleus, raphe nuclei, and other brainstem and hypothalamic nuclei. They innervate the cerebellar and vestibular nuclei and all cortical layers, with some preference from the molecular layer. Modulatory inputs synapse onto Purkinje cells as well. Afferents from the locus ceruleus release norepinephrine, those from the raphe release serotonin, those from the brainstem release acetylcholine, and those from the hypothalamus release histamine. Synapses of nonlaminar afferents have been difficult to define at the electron microscopic level because of their sparseness and lack of distinctive structural characteristics. As a result, little is known about the specifics of their connectivity or their functional roles in modulating cerebellar activity (Ito, 2009).

Development

Cerebellar cell types develop at different times and at different places. First to be formed are the neurons of the cerebellar nuclei. They are followed soon after by Purkinje cells, which originate in the ventricular epithelium and migrate to their ultimate location in the cortex. Golgi cells, basket cells, stellate cells, astrocytes, and Bergmann glia also originate at the ventricular epithelium and migrate to their final positions after the migration of the Purkinje cells. The Bergmann glia later guide the descent of the granule cells from the external granule cell layer.

Once the Purkinje, Golgi, basket, and stellate cells have formed, climbing fibers enter the cerebellum from the inferior olive and begin to innervate the Purkinje cells. Much later, after the Purkinje cells have begun to receive synapses from parallel fibers, most of the climbing fiber contacts with Purkinje cells are eliminated. Mossy fibers also enter the cerebellum and grow to the level just below the Purkinje cell layer. They will ultimately synapse on granule cells, which have yet to arrive.

The granule cells first develop at a very distant site, posterior and lateral in the anlagen at the rhombic lip. They then crawl over the Purkinje cells to form an external granule cell layer. Some granule cells may cross the midline to the other side of the cerebellum as they make this migration. The granule cells extend their axons, which branch to form parallel fibers that run as coronal beams through the dendrites of the Purkinje cells. Only then do the granule cells descend, guided by the Bergmann glia, to form the internal granule cell layer. Each granule cell is connected by the proximal portion of its axon to the branch point of its parallel fibers, which remain in the external granule cell layer. In the internal granule cell layer, each granule cell dendrite is contacted by the terminals of the mossy fibers and gradually forms synaptic connections with them. Synapses are also made between granule cells and the intrinsic cortical inhibitory neurons—the Golgi, stellate, and basket cells. Notably, the output side of the cerebellar circuit (vestibular and nuclear neurons and Purkinje cells) forms first, the input side (mossy fibers) then arrives, and the "matrix" that connects the two (the granule cells and intrinsic inhibitory neurons) is the last to develop (Ray & Dymecki, 2009; Sillitoe & Joyner, 2007).

In humans, the first cerebellar structures develop at approximately 32 days after fertilization, and development is not completed until after birth. The cerebellar cortex of the early embryo has six distinct layers, but this number is ultimately reduced to three. The cortex begins to differentiate slightly earlier in the vermis, flocculus, and median sections of the hemispheres than in the lateral hemispheres. By 7 months after fertilization, the cerebellar nuclei have attained the shape and location they will have in the adult. At birth, the cerebellar cortex consists of four uneven layers, and the Purkinje cells and basket cells are weakly developed. The fourth layer (the external granule cell layer) disappears within the first postnatal year. In humans, full myelination of cerebellar connections is not complete until the second year of life (Brody, Kinney, Kloman, & Gilles, 1987; Wang and Zoghbi, 2001).

Summary

The cerebellum is a prominent structure in the nervous systems of most vertebrates. The cerebellum receives inputs from many parts of the nervous system but predominantly influences motor circuits and the cerebral cortex. This suggests that many different kinds of information are brought together to coordinate and adaptively control movement and certain cognitive functions. The intrinsic structure is relatively simple and very stereotyped throughout the cerebellum, very different from that of the cerebrum. The ontogenetic development is conspicuously late: cell migrations and fiber connections continue to occur after birth and the development of the rest of the motor nervous system.

CELLULAR AND SYNAPTIC PERFORMANCE OF THE CEREBELLAR CIRCUIT

Distinctive Properties of Cerebellar Neurons

The continuous high firing rates of both Purkinje cells and vestibular/nuclear neurons are unusual in comparison to most other brain regions. These cerebellar neurons

are not only active during behaviors, but they also play critical roles in the induction and expression of long-term memories. These constraints pose a distinct set of challenges for the underlying cellular and synaptic physiology. This section examines the adaptations and performance of several aspects of the cerebellar circuit: high spontaneous firing rates, four distinctive synaptic connections, and multiple forms of synaptic and intrinsic plasticity.

Spontaneous, Continuous High Firing Rates

The advantage of the high firing rates of Purkinje and nuclear/vestibular neurons is thought to be the resulting rapid, bidirectional modulation of firing rate that enables tight temporal control over movement and planning. However, high firing rates incur a significant metabolic cost. Each time a neuron fires an action potential, sodium and calcium enter the neuron and potassium exits. These ions must be pumped back across the membrane with ATP-driven ion exchangers to regain their balance. Purkinje and cerebellar/vestibular nucleus cells express a specific complement of ion channels that help minimize this cost and maintain the cell's spontaneous rate. They are enriched for Kv3 potassium channel subtypes, which have rapid kinetics that speed cell repolarization. As a result, action potentials are brief in duration, limiting the window for sodium and calcium entry, so that fewer of these ions will have to be pumped back outside. Rapid repolarization together with specialized sodium channels (Bant & Raman, 2010; Raman & Bean, 1991) minimizes the inactivation of sodium channels that typically occurs during each action potential (Carter & Bean, 2011; Gittis et al., 2010). These biophysical properties are critical for enabling cerebellar neurons to rapidly and precisely modify their downstream targets.

Mossy Fiber Glomerulus

The mossy fiber axon terminates in glomerular structures in the granule cell layer, each of which can contain a dendritic terminus from each of ~20 to 30 granule cells and one Golgi cell basal dendrite. Each mossy fiber forms 15 to 50 such glomeruli per folium, so that it directly influences as many as 1000 granule cells and 50 Golgi cells in one folium. Mossy fiber firing patterns vary across the cerebellum, depending on the source of the fiber, and can range from prolonged firing at tens of Hz to short bursts at instantaneous frequencies upwards of 200 Hz. The mossy fiber to granule cell synapse achieves reliable transmission of high bursts by rapid recycling and reloading of a large pool of available vesicles (Saviane & Silver, 2006). It is thought that the conjunctive activity of several mossy fiber inputs is required to trigger granule cell firing *in vivo*, but it is possible that just one is sufficient. The small size and close quarters of granule cells make it technically challenging to record the spiking of individual neurons, leaving this question open to further study.

Climbing Fiber Synapse

The climbing fiber that arises in the inferior olive makes an intricate, ivy-like connection with each of its Purkinje cell targets, synapsing along the entire dendritic arbor at ~2–3 μm intervals. Ascending axonal branches also have contact points with molecular layer interneurons, although the dominant mode of transmission appears to be through spillover of glutamate from nearby release sites (Szapiro & Barbour, 2009). While the climbing fiber ascending synapse onto Purkinje cells is the best studied, the axon also makes a number of transverse branches of unknown function, with few conventional synapses (Nishiyama, Fukaya, Watanabe, & Linden, 2007).

Each synaptic contact between the climbing fiber and Purkinje cell is ensheathed by glial processes, deriving from Bergmann glia, that reduce the spillover of glutamate from the synaptic cleft to other sites. This specialization may be due to the large amount of glutamate released by the climbing fiber: not only does each axon make hundreds of contacts with one Purkinje cell, each contact is able to release multiple vesicles of glutamate at a time (Wadiche & Jahr, 2001). Furthermore, inferior olive neurons can fire complex spikes as well, in which a high-frequency burst (3–4 spikes at 100–200 Hz) is successfully transmitted down the axon, leading to an even larger glutamate release (Mathy et al., 2009). The result of all this glutamate release is the Purkinje cell complex spike. Each complex spike causes ~2 to 3 spikes to propagate down the Purkinje cell axon, at instantaneous rates of up to ~400 Hz (Fig. 31.5). However, the specific waveform of the complex spike, and thus the number of action potentials that propagate down the Purkinje cell axon, is determined by the characteristics of climbing fiber synaptic transmission and is subject to change. Interestingly, the ~50 Hz simple spike activity in a Purkinje cell is often suppressed for tens of ms after the complex spike, so that the net effect of this synapse is a burst-pause. The flood of calcium triggered by climbing fiber activity is a critical signal for parallel fiber plasticity (see below).

Purkinje Cell to Cerebellar Nuclear Neuron Synapse

Purkinje cells fire simple spikes spontaneously at ~50 Hz, modulate their firing up to 100–200 Hz, and can achieve instantaneous frequencies of up to 400 Hz in response to climbing fiber input. How do the output neurons of the cerebellar circuit respond? The Purkinje cell synapse onto cerebellar nucleus neurons is robust. Each axon terminal is large (~3–4 μm diameter) and contains around 10 individual release sites (Telgkamp, Padgett, Ledoux, Woolley, & Raman, 2004). Furthermore, each

Purkinje cell axon makes several synaptic contacts onto a single cerebellar nucleus neuron, providing a total of 100 to 200 release sites at a unitary connection. Although the probability of transmitter release is low at each site, the sheer number of release sites guarantees that a given Purkinje cell can reliably release transmitter to a cerebellar nucleus neuron even at high rates of presynaptic firing.

The high spontaneous rates of cerebellar nucleus neuron firing enable Purkinje cell inhibition to sculpt postsynaptic responses in several different ways. Dozens of Purkinje cell axons converge onto individual nuclear neurons. Synchrony in the firing of converging Purkinje cells enables inhibition to produce precise delays in postsynaptic firing with little effect on overall rate; conversely, asynchronous firing of Purkinje cells results in a reduction in nuclear neuron firing rates (Person & Raman, 2012). Most nuclear neurons respond to the *offset* of hyperpolarization with a transient increase in firing rate, termed *rebound firing*, before settling into a new, elevated rate relative to the inhibited state. Whether Purkinje cell inhibition is sufficiently strong to engage rebound firing remains unclear. One working hypothesis is that the relief of Purkinje cell inhibition, evoked by a synchronized pause in Purkinje cell firing, serves as a precisely timed signal to increase cerebellar nuclear neuron firing and modulate downstream behavior. In this scenario, the timing of the Purkinje cell pause would be controlled by temporal coding provided by granule cell inputs so that learned Purkinje cell responses can be temporally specific.

Cerebellar Nuclear to Olivary Input

GABAergic neurons in the cerebellar and vestibular nuclei give rise to thin axons that project to the inferior olive, closing the feedback loop of the cerebellum. Little is known about the firing rates of these GABAergic neurons with respect to behavior, because of the difficulty in identifying these neurons during *in vivo* recordings. However, the connection is clearly required for extinction of learned reflexes (as described below). Nuclear and vestibular neurons make synapses on olivary dendrites. *In vivo* studies have shown that stimulation of the nucleo-olivary tract causes a slow, long-lasting inhibition of olivary firing. Recent work has revealed that the sluggishness of this response derives from asynchronous release of GABA, yielding a barrage of miniature synaptic events. The resulting synaptic current is small in amplitude but lasts tens of milliseconds (Best & Regehr, 2009). Interestingly, inputs from vestibular nuclei to the olive do not share this characteristic: they are phasic and rapid in time course (Box 31.2). Nuclear synapses onto olivary dendrites occur preferentially close to the gap junctions that interconnect olivary neurons (De Zeeuw et al., 1998). This localization suggests that nuclear input might selectively weaken olivary coupling (Lang, Sugihara, & Llinas, 1996).

Long-Term Depression and Potentiation at the Parallel Fiber to Purkinje Cell Synapse

The cerebellum is the site of certain types of memory formation and storage. Activity-dependent plasticity has been found in synapses and neurons throughout the cerebellar circuit (Carey, 2011). The following section describes two forms of cerebellar plasticity that have been long hypothesized to mediate cerebellar-dependent learning.

Marr first proposed the existence of plasticity at the parallel fiber to Purkinje synapses whose induction is controlled by the climbing fiber input (Marr, 1969). Later, Masao Ito and colleagues demonstrated climbing fiber-controlled induction of long-term depression (LTD) at these synapses (Ito, Sakurai, & Tongroach, 1982; Ito, 1996). This plasticity and the related implication that the climbing fibers provide a "teaching" input to the cerebellum are at the core of theories of cerebellum function (Ito, 2001).

The strength of the granule cell to Purkinje cell synapse can be studied over the period of several hours in slice preparation (or more rarely, *in vivo*). Classic early studies showed that when a set of parallel fibers is stimulated repeatedly and in conjunction with the other major input to a Purkinje cell, the climbing fiber, the amplitude of the parallel fiber synapse is reduced. This depression is maintained for the duration of recording and thus is termed *parallel fiber long-term depression* (Fig. 31.6A). Somatic depolarization of the Purkinje cell can substitute for climbing fiber activity. Subsequent studies revealed more about the relative timing required between parallel fiber and climbing fiber activation to induce parallel fiber LTD: best induction of LTD requires parallel fiber activation 100–250 ms prior to the climbing fiber input. This agrees well with *in vivo* studies suggesting that climbing fiber inputs modify synapses onto Purkinje cells that were active 100–200 ms prior (Raymond & Lisberger, 1996). In contrast, repeated stimulation of the parallel fiber pathway by itself, in the absence of climbing fiber activity, leads to potentiation of the stimulated inputs (*long-term potentiation*) (Fig. 31.6B) (Coesmans, Weber, De Zeeuw, & Hansel, 2004). Mechanistic analyses reveal that paired parallel and climbing fiber input produces an mGluR-driven signaling cascade in Purkinje cells (Linden et al., 1991), leading to the removal of AMPA-type glutamate receptors from the parallel fiber synapse (Wang & Linden, 2000). However, the requirement of parallel fiber LTD for cerebellar motor learning is unclear. Using genetically modified mice in which the mechanisms responsible for the expression of LTD are specifically impaired, some authors have found impairments in motor learning while others see no changes (De Zeeuw et al., 1998; Schonewille et al., 2011).

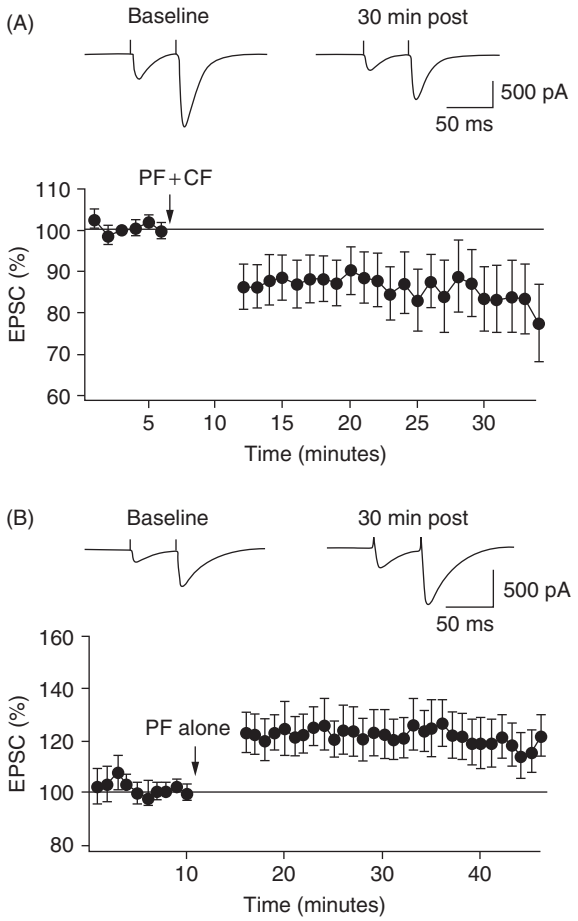


FIGURE 31.6 Whole-cell recordings from Purkinje cells showing the synaptic result of parallel fiber stimulation over time. (A) After a baseline period, parallel fiber stimulation is paired with conjunctive climbing fiber stimulation every 1 s for 5 min (arrow). Afterwards, the parallel fiber synaptic amplitude is depressed (LTD). (B) The parallel fiber is stimulated as in A, but in the absence of climbing fiber input potentiated (LTP). *Adapted from Coesmans et al. (2004).*

Synaptic and Intrinsic Forms of Plasticity in Vestibular and Cerebellar Nucleus Neurons

Miles and Lisberger proposed the hypothesis that cerebellar learning is mediated by plasticity downstream of Purkinje cells, at synapses onto cerebellar and vestibular nucleus neurons (Miles & Lisberger, 1981). They proposed that an "error signal" is calculated in the cerebellar cortex and the result of that calculation is conveyed by the Purkinje cells to their target neurons, where it serves as a teaching input to control the induction of plasticity. *In vivo* recording studies provided strong evidence for this hypothesis during adaptation of the VOR and during eyelid conditioning.

Recent work has found two major forms of plasticity in Purkinje cell target neurons. The first is changes in intrinsic excitability. Target neurons in the vestibular and cerebellar nuclei can modify their intrinsic input-output functions, so that a synaptic input of a given size that once elicited a modest response (as evaluated by postsynaptic spiking) now elicits a strong one, or vice versa (Shadmehr, Smith, & Krakauer, 2010; Zhang & Linden, 2003). This type of plasticity might be useful for generalized changes of behavioral gains—the magnitude of the response to a given stimulus.

The second type of plasticity is synaptic. The strength of synapses from excitatory mossy fibers onto their postsynaptic targets in the cerebellar or vestibular nuclei can be bidirectionally modified. The learning rules, inspired by ideas from Lisberger and Miles applied to eyelid conditioning (Mauk & Donegan, 1997), are distinctive. Long-term potentiation (LTP) and LTD of the mossy fiber to nuclear neuron synapse are controlled by Purkinje cell inhibition in conjunction with afferent synaptic stimulation. Afferent input combined with release from hyperpolarization is required to trigger an increase in synaptic strength, whereas afferent input without release from hyperpolarization triggers decreases in synaptic strength (McElvain, Bagnall, Sakatos, & du Lac, 2010; Pugh & Raman, 2009; Zhang & Linden, 2003). Increases and decreases in Purkinje cell firing rate are the most obvious potential sources of cerebellar nuclear hyperpolarization and release, respectively. Each of these forms of plasticity is a candidate substrate of cerebellar learning, and their relative roles are under active debate and investigation.

CEREBELLAR FUNCTION

The stereotyped circuitry and cellular mechanisms of signaling and plasticity within the cerebellar cortex and nuclei suggest a common role for diverse regions of the cerebellum. But what exactly is that role? Dysfunction of the cerebellum leads to a variety of clinical manifestations (see Box 31.3). As described in the following sections, the cerebellum is thought to adaptively modify downstream circuits, optimizing performance in the face of changes in the external and internal environments. To optimize circuit function and behaviors effectively, the cerebellum requires information from sensory, motor, and cortical systems. Each region of the cerebellum exerts its influence via cerebellar or vestibular nucleus neurons that themselves influence a variety of downstream circuits and behaviors.

Mossy Fibers and Climbing Fibers Convey Distinct Types of Information to Cerebellar Circuits

Precerebellar mossy fiber neurons distributed throughout the midbrain (red nucleus), pons, medulla, and spinal cord convey signals that inform the cerebellum about the

BOX 31.3

CLINICAL TESTING CAN REVEAL CEREBELLAR DAMAGE

In the early 1900s, Gordon Holmes carefully described the movement deficits associated with discrete cerebellar lesions caused by gunshot wounds (Holmes, 1939). Lesions caused by cerebellar infarct also provide information about the role of portions of the cerebellum, although interpretation of these data can be difficult because of the frequent spread of damage to extracerebellar regions. The common theme through these studies is that cerebellar damage does not impair sensory perception, but instead causes deficits in motor coordination and adaptation. These deficits are consistent with a role for the cerebellum in producing a *forward model* of movement, which predicts the expected sensory result of a movement and updates the forward model if errors occur (Shadmehr & Krakauer, 2008). Several types of cerebellar deficits are seen in human patients:

Ataxia is a condition that involves lack of coordination between movements of body parts. The term often is used in reference to gait or movement of a specific body part, as in "ataxic arm movements." Ataxia of posture, stance, and rhythm can derive from damage to vestibular and fastigial nuclei and vestibulocerebellum, while the intermediate and dentate nuclei and associated cortex appear more responsible for limb placement and multisegmental coordination. **Asynergia** is an inability of cerebellar patients to correct their movements for the interaction torques between multiple joints, leading to inaccurate reaching or stepping movements.

Dysmetria is an inability to make a movement of the appropriate distance. Hypometria is undershooting a target, and hypermetria is overshooting a target. Patients with cerebellar damage tend to make hypermetric movements when they move rapidly and hypometric movements when they move more slowly and wish to be accurate.

Dysdiadochokinesia is an inability to make rapidly alternating movements of a limb. It appears to reflect abnormal agonist-antagonist control. It is thought to be due to damage to the interpositus and/or dentate nuclei and resulting perturbation of muscle stretch reflexes.

Action tremor, or **intention tremor**, is an involuntary oscillation that occurs during limb movement and disappears when the limb is at rest. Cerebellar action tremor is generally of high amplitude and low frequency (3–5 Hz). **Titubation** is a tremor of the entire trunk during stance and gait. Lesions of cerebellar target structures (e.g., the red nucleus and the thalamus) often result in cerebellar outflow tremor, or postural tremor. Most prominent when a limb is actively held in a static posture, postural tremor attenuates during limb movement and disappears when the limb is at rest. Like dysdiadochokinesia, these tremors are also thought to be caused by deficient stretch reflexes.

Hypotonia is an abnormally decreased muscle tone. It is manifest as a decreased resistance to passive movement, so that a limb swings freely upon external perturbation. Hypotonia often is limited to the acute phase of cerebellar disease.

Nystagmus is an involuntary and rhythmic eye movement that usually consists of a slow and a fast phase. In a unilateral cerebellar lesion, the fast phase of nystagmus is toward the side of the lesion.

Patients with cerebellar damage are impaired both in normal motion and during motor adaptation. For example, a normal subject can adapt to wearing prism eyeglasses that shift the world to one side. After the prisms are removed, the subject shows a slow return to normal behavior, indicating that learning took place. In contrast, a cerebellar patient shows neither adaptation to the prisms, nor any sign of learning afterwards. Saccade learning is similarly impaired in cerebellar patients.

There is some promise for rehabilitation of cerebellar patients. Recent work suggests that these patients can learn to correct for small errors, either by using remaining functional circuits or through other brain circuits entirely. Cumulative learning of small adjustments can produce an overall large change in the resulting behavior, suggesting that compensation may be possible for certain types of deficit (Bastian, 2011).

Martha Bagnall and Tom Thach

sensory world, movements, and cortical plans for action (including, perhaps, cognitive plans that do not result in overt movements). Each type of mossy fiber neuron conveys a specific type of information to the cerebellum (Kolkman, McElvain, & du Lac, 2011). Neurons in the vestibular and external cuneate nucleus and in Clarke's column in the spinal cord transmit sensory signals to the cerebellum (vestibular, somatosensory, and proprioceptive). Sensory information about internal state and autonomic function are handled by neurons in the lateral reticular nucleus. "Efference copy" signals (which inform the brain about commands for movement sent to motor neurons) are conveyed to the cerebellum by neurons in the nucleus prepositus and pontine reticular formation. Neurons in the pons mediate the connections between the cerebral cortex and the cerebellum. Some mossy fiber neurons fire

high-frequency bursts of action potentials, while others fire tonically and modulate their firing rates in response to inputs, thus providing the cerebellum with continuous streams of information.

In contrast, neurons in the inferior olive fire rarely and convey discrete signals to the cerebellum. Each of several subdivisions of the inferior olive comprises neurons that encode unexpected signals, or errors, in a particular domain. For example, neurons in the caudal portion of the dorsal cap of the olive increase their firing rates (from 1 to 2 Hz) under only one condition: when images move horizontally across the retina. Neighboring neurons in the rostral dorsal cap fire when images move vertically. Some olivary neurons are activated in response to head motion and somatosensory, proprioceptive, nociceptive, or viscerosensory signals, while others appear to be activated when movements are planned or initiated by the superior colliculus or cerebral cortex.

Cerebellar and Vestibular Nuclei Influence Several Neural Circuits and Behaviors

Vestibular Nucleus

The vestibular nuclei comprise several subdivisions (see Chapter 29) and are important for several different behaviors. The vestibulo-ocular reflex (VOR) is the simplest and most extensively studied behavior under cerebellar control. The VOR enables clear vision during head movements by producing compensatory eye movements (see Chapter 32). A simple form of motor learning—adaptation—occurs whenever images move persistently during head movement. Adaptation can either increase or decrease the gain of reflexive eye movements. The flocular lobe of the cerebellum is required to induce adaptive changes in the VOR. Remarkably, although the VOR is mediated by as few as two muscles, Purkinje cells in the flocculus synapse onto at least five different cell types in the vestibular nucleus (Shin et al., 2011). The relative contribution of plasticity in the flocculus versus these vestibular nucleus neurons to motor learning in the VOR remains to be elucidated.

Cerebellar inputs to lateral vestibular nucleus are critical for coordinating and calibrating postural movements (see Chapter 29). Caudal vestibular nucleus neurons, which convey information about head tilt and gravity to autonomic brainstem regions (Holstein et al., 2011), are responsible for increasing blood pressure rapidly during the transition from lying down to standing up.

Fastigius (Medial Cerebellar Nucleus)

The fastigius controls all musculature involved in stance and gait. It receives input from the vermal cortex, vestibular complex, lateral reticular nucleus, and (indirectly) spinocerebellar pathways. Ablation of the fastigius dramatically impairs movements requiring control of equilibrium, such as unsupported sitting, stance, and gait. Single-unit recordings in the fastigius and vermal cortex of decerebrate cats, but not the interpositus or dentate, have shown neural discharge that is correlated with both walking and scratching movements. Rostral fastigial neurons are critical for the regulation of blood pressure and other autonomic functions. The caudal portion of the fastigius contains neurons that project to midline nuclei of the thalamus (Steriade, 1995), as well as neurons responsible for maintaining the accuracy of saccadic eye movements.

Interpositus (Intermediate Cerebellar Nucleus)

Inactivation of both the interpositus and the dentate with cooling probes elicits tremor that depends on proprioceptive feedback but is also influenced by vision. Inactivation of the interpositus causes a large-amplitude, 3- to 5-Hz action tremor as animals reach for food but has minimal effects on gait. Neurons in the interpositus fire when the holding position of a limb is perturbed. Their activity appears to control antagonist muscles that check the reflex movement of the limb to its position. Interpositus neurons also modulate their activity in relation to sensory feedback, including that from tremor accompanying movement, consistent with its role in controlling the antagonist muscles to dampen the tremor. Other evidence suggests that the interpositus is important in determining whether the pattern of activity in muscles acting at a joint represents reciprocal activation or co-contraction. During behaviors that involve co-contraction, interpositus neurons fire as if activating both agonist and antagonist muscles, and Purkinje cells are silent. In behaviors where agonists and antagonists are reciprocally active, both interpositus cells and Purkinje cells fire in similar patterns. One interpretation of these results is that alternating firing in Purkinje cells creates (through inhibition) a similar pattern of activity in the interpositus cells, which in turn produces the alternation between agonist and antagonist muscles. Other work suggests that the interpositus contributes to stretch reflex excitability by controlling the discharge of gamma motor neurons (Chapter 28). The interpositus also houses neurons required for classical conditioning of the eyelid response, as described below.

Dentate (Lateral Cerebellar Nucleus)

Neuronal activity in the dentate precedes the onset of movement and may also precede firing in the motor cortex. Dentate neurons fire preferentially at the onset of movements that are triggered by mental associations. Single-unit recordings in the motor cortex, dentate, and interpositus can be correlated with electromyograms as monkeys make wrist movements in response to stimuli. In tasks where movements are triggered by light, the order of activity is dentate, motor cortex, interpositus, and muscles. When a transient force perturbs the wrist, however, the firing order is muscles, interpositus, motor cortex,

and dentate. These results suggest that the dentate influences movements triggered by stimuli that are mentally associated with the movement, whereas the interpositus is more involved in compensatory or corrective movements initiated via feedback from the movement itself. In other studies, the dentate responded strongly when movements were triggered by either visual or auditory signals but not proprioceptive signals. Firing in both dentate and interpositus neurons is thought to relate more closely to movements involving multiple joints than to those involving single joints. Neither nucleus codes for any specific parameter (e.g., velocity, amplitude, or duration) during single-jointed movements. In sum, the dentate plays a role in initiating movements that require a mental interpretation of the visual or auditory signal, and both the dentate and the interpositus are increasingly active during multijointed movements.

Theories of Cerebellar Function

With the knowledge available about cerebellar anatomy, cellular and synaptic physiology, interconnectedness with the rest of the brain, and effects of cerebellar damage on behavior, research has been strongly influenced by a variety of computationally oriented ideas about cerebellar function. Pronounced deficits in muscle tone following cerebellar damage led to early ideas that the cerebellum could serve to reinforce tone. Other theories emphasized the importance of the cerebellum in motor coordination. The organization of parallel fibers and their relatively slow conduction times were incorporated into an influential model, proposed by Braitenberg and colleagues, which postulated that the cerebellar circuit produces precise timing. Behavioral experiments in patients with cerebellar damage have lent support to the notion that the cerebellum is critical for specifying motor timing (Spencer, Zelaznik, Diedrichsen, & Ivry, 2003). The 10-Hz tremor produced by damaging inferior olive neurons with the drug harmaline and observations that olivary neurons can emit 8- to 10-Hz bursts of firing led Llinas, Welsh, and others to suggest that the cerebellum provides clock-like timing signals to the rest of the brain.

The wealth of sensory and cortical information conveyed to the cerebellum, the indirect influence of cerebellar output nuclei on motor neurons, and the extensive interconnection to nonmotor regions of the cerebral cortex (Strick, Dum, & Fiez, 2009) suggest that the primary role of the cerebellum may not be motor *per se* but instead may be to evaluate sensory or cortical information in the context of ongoing movements or plans. The hypothesis that the cerebellum predicts the consequences of movements is supported by behavioral studies in patients with cerebellar damage (Bastian, 2011; Shadmehr, Smith, & Krakauer, 2010). Cerebellar involvement in certain aspects of cognitive and emotional processes is consistent with a prominent role for prediction in cerebellar function (Box 31.4).

BOX 31.4

COGNITION, EMOTION, AND THE CEREBELLUM

It no longer is viable to consider the function of the cerebellum as being confined to the control of voluntary movement, speech, and equilibrium. Considerable evidence suggests that the cerebellum is critical also for cognition and emotion. Anatomical studies demonstrate that the cerebellum is an important part of the distributed neural circuitry that subserves cognitive processing. The frontal, association, and paralimbic cerebral cortices known to subserve higher-order functions are linked with the cerebellum in a precisely organized system of feedforward and feedback loops, and physiological studies indicate that these pathways are functionally relevant (Strick et al., 2009). Cerebellar ablation and stimulation experiments in animals have demonstrated cerebellar influences on many nonmotor functions, including classical conditioning, navigational skills, cognitive flexibility, sham range, predatory attack, and aggression. Functional neuroimaging investigations of the morphologic correlates of cognitive processing using PET and fMRI in humans have revealed sites of activation in the cerebellum in a number of cognitive tasks. These include linguistic processing, verbal working memory, shifting attention, mental imagery, classical conditioning, motor learning, sensory processing, and modulation of emotion. Furthermore, there appears to be a topographic organization of the sites within the cerebellum activated by these different cognitive processes.

Clinical investigations of adults and children with diseases confined to the cerebellum have defined a cerebellar cognitive affective syndrome characterized by impairments of executive processing, working memory, visual spatial reasoning, language disturbances (ranging from mutism to agrammatism), and a flattened or inappropriate affect. The net effect of these deficits is a lowering of overall intellectual ability. The posterior lobe of the cerebellum appears particularly important in the generation of this syndrome, and the vermis is consistently involved when the affective component is pronounced. Elements of this clinical syndrome have also been noted in patients with developmental cerebellar

BOX 31.4 (*cont'd*)

anomalies, cerebellar degeneration, autism, and fragile X syndrome. In addition, cerebellar abnormalities, particularly in the vermis, have been observed in patients with schizophrenia.

The relationship between the cerebellum and nonmotor function has been conceptualized as follows.

1. The cerebellum is able to subserve cognitive and emotional functions because it is anatomically interconnected with the associative and paralimbic cortices.
2. The convergence of inputs from multiple regions of the cerebral cortex to adjacent areas within the cerebellum facilitates cerebellar regulation of supramodal functions.
3. The cerebellar contribution to cognitive and emotional processes is one of modulation rather than generation.
4. The cerebellum performs computations for cognitive functions similar to those for the sensorimotor system—but the nature of the information being modulated is different.
5. The disruption of the cerebellar influences on higher functions can lead to dysmetria of thought or impairment of mental agility.
6. The cerebellum may play an important role in the development of cognitive and emotional function.

The potential for further discovery in this field places the cerebellum, previously thought of as a motor control device, in the forefront of current behavioral neuroscience research.

Sascha du Lac and Jeremy D. Schmahmann

The most influential theory, which implicated the cerebellum in motor learning, was proposed more than four decades ago by Marr (1969). At that time, the basic cerebellar circuit had been recently identified; from it, Marr inferred the essential learning attributes of the cerebellum including plasticity at the granule cell to Purkinje cell synapses and the role of climbing fibers as a teaching input to control the induction of this plasticity. These ideas inspired Masao Ito's demonstration, in the context of the VOR, of climbing fiber controlled LTD (and much later the demonstration of the induction of LTP when granule to Purkinje synapses are active without a climbing fiber input; Fig. 31.6). An early link between circuit and behavior came from experiments showing changes in Purkinje cell simple and complex spikes during the course of motor learning (Fig. 31.7). The importance of learning in the cerebellum and how it may contribute to cerebellar computation and function remain actively investigated issues.

Computation as a Unifying Theme

What computations does the cerebellum perform on its inputs? Like all brain systems, the essence of what the cerebellum does is this: it receives inputs from other parts of the brain, acts on these inputs with rules determined by its connectivity and synaptic and cellular properties, and produces output to other brain systems. Given that the synaptic organization of the cerebellum is relatively uniform, it is likely that the same computational principles apply qualitatively for cerebellar regions involved in motor, autonomic, and cognitive functions.

Analyses of what (and how) the cerebellum computes have been facilitated greatly by the advantages provided by a limited number of experimental preparations amenable to quantitative behavioral analyses. These include adaptation of the vestibulo-ocular reflex, saccadic eye movements, and reaching, as well as motor learning in smooth pursuit eye movements. Insights from these behaviors have been complemented by the analysis of classical associative (Pavlovian) conditioning of eyelid responses.

Cerebellum and Eyelid Conditioning

The procedures of eyelid conditioning are relatively straightforward: an animal is presented with training trials that involve a relatively neutral stimulus such as a tone paired in time with a reinforcing "unconditioned stimulus" such as a puff of air or mild electrical stimulation near the eye; we will use "puff" for convenience. In an untrained subject the puff elicits a reflex eye blink response, but the tone elicits no response. After a sufficient number of training trials, the tone comes to elicit learned eyelid responses (Fig. 31.8A; sample eyelid sweeps). Remarkably, the timing of these learned eyelid responses precisely anticipates the timing of the puff onset: animals can be trained using a range of temporal intervals between tone and air puff onsets and are capable of learning responses appropriate for each interval: the learned responses peak just before the air puff occurs (Fig. 31.8B). This learning is bidirectional: presenting the tone without the air puff to an already trained animals results in extinction of the learned eyelid responses (Fig. 31.8C).

Support for the conclusion that eyelid conditioning is mediated by cerebellar learning comes from lesion studies and from experiments employing electrical stimulation and reversible pharmacological inactivation. Lesions of either the cerebellar cortex or the cerebellar nucleus

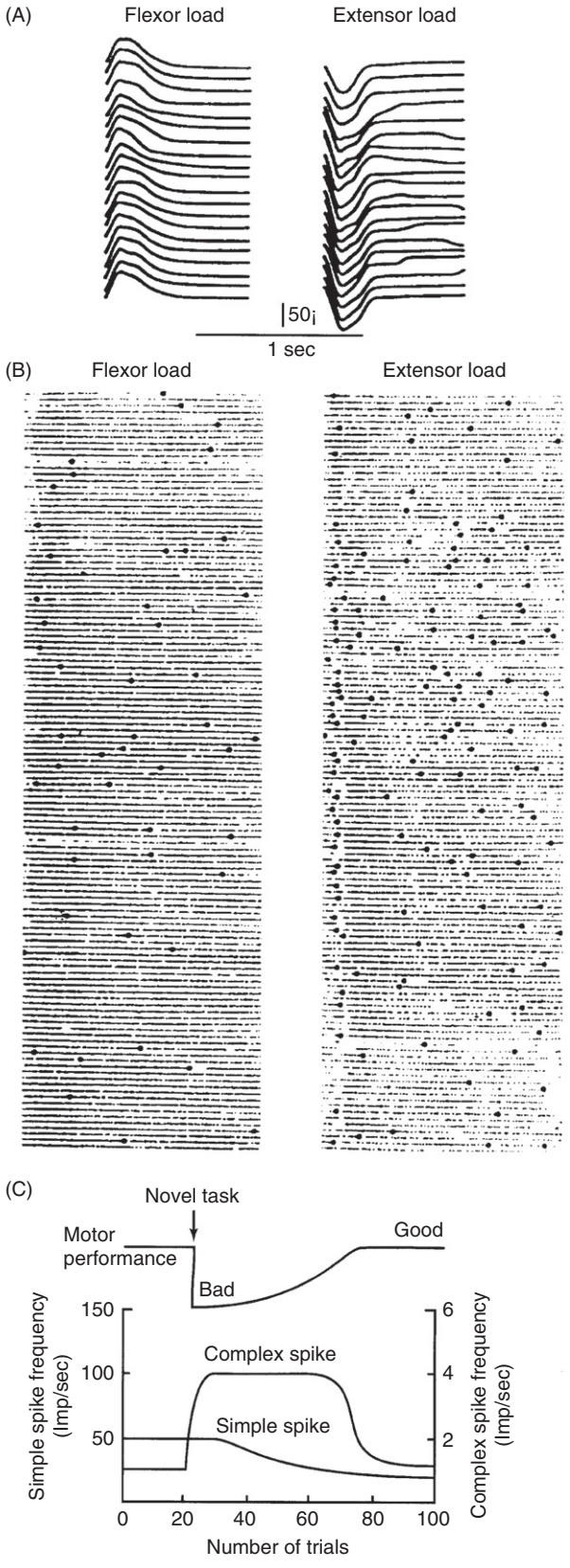


FIGURE 31.7 Simple and complex spikes recorded during motor learning task. *Adapted from Ito; original reference Gibert & Thach (1977).*

(anterior interpositus) prevent naive animals from acquiring learned eyelid responses (Garcia, Steele, & Mauk, 1999). Robust eyelid conditioning can be induced by direct electrical stimulation of mossy fibers, which can substitute for the tone, combined with stimulation of climbing fibers, which can substitute for the air puff. Reversible inactivation studies provide complementary evidence for a critical role of the cerebellum: pharmacological inactivation of the red nucleus (which receives inputs from the interpositus cerebellar nucleus) during training precludes expression of responses, but learned responses are present in the first trials after removal of the block. This indicates that the plasticity responsible for learning happens upstream of the red nucleus. Such studies demonstrate that eyelid conditioning is mediated by neuronal plasticity within the cerebellar cortex, the cerebellar nuclei, or both (Fig. 31.9).

Behavioral analyses suggest differential roles for the cerebellar cortex and nuclei in eyelid conditioning. Lesions to the cerebellar cortex disrupt the learned timing of the conditioned responses. In intact animals, learned responses peak at the time that the air puff was presented. In contrast, animals with cerebellar cortex lesions are unable to learn appropriately timed responses and instead respond at a short, fixed latency. Subsequent work using stimulation and reversible inactivation studies has shown that these short-latency responses are mediated by plasticity in the cerebellar nucleus. Together, these findings suggest (1) cerebellar learning is mediated by plasticity in both the cerebellar cortex and the cerebellar nuclei, (2) the cerebellar cortex is necessary for the learned and adaptive timing of the conditioned responses, and (3) consistent with the Miles and Lisberger theory, the cerebellar cortex provides the teaching signal for the induction of plasticity in the cerebellar nuclei.

Eyelid conditioning studies have revealed a role for the inhibitory feedback from cerebellar nuclei to the inferior olive (Medina, Nores, & Mauk, 2002). This portion of the cerebellar circuit has been particularly difficult to examine because olivary-projecting neurons are intermingled with other cerebellar nucleus neurons and because the olive itself is a tiny structure located deep in the brainstem. Pharmacological blockade of synaptic receptors implicates the inferior olive in memory storage. Learned eyelid responses are extinguished (a form of forgetting) if the tone conditioned stimulus is presented repeatedly by itself, without the air puff. Eventually, animals cease to emit appropriately timed eyelid responses to the tone. In previously trained animals, infusion of GABAergic synaptic blockers into the olive during tone-alone training completely prevents extinction. Conversely, infusion of excitatory synaptic antagonists causes extinction, even when the animals are being presented with the normal training stimulus of tone together with the air puff.

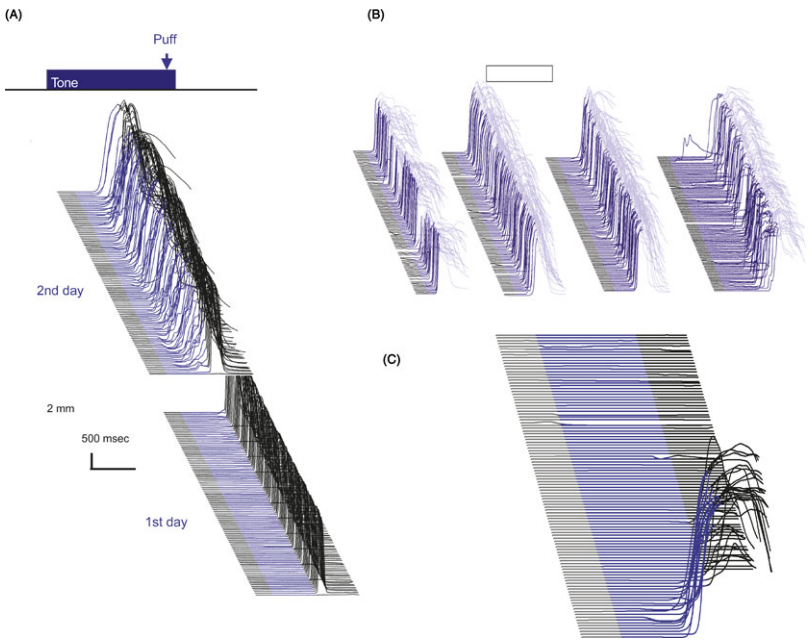


FIGURE 31.8 Eyelid conditioning. (A) Top, schematic of the training procedure. A tone is presented for a brief duration (usually ~500 ms), coterminating with a brief air puff to the eye. Bottom, traces of eyelid movement during training. Each line represents one trial; the period in blue is the tone presentation. On the first day of training, animals only blink in response to the air puff; by the second day, the eyelid begins to move before the end of the tone, eventually reaching peak movement that is well timed to block the air puff. (B) Animals can learn a range of tone durations, with appropriately timed blinks regardless of the interval between tone onset and air puff. (C) Extinction of prior learning. A well-trained animal blinks at the appropriate time even when the air puff is not presented, but after repeated trials with no air puff, the animal ceases to blink in response to the tone alone.

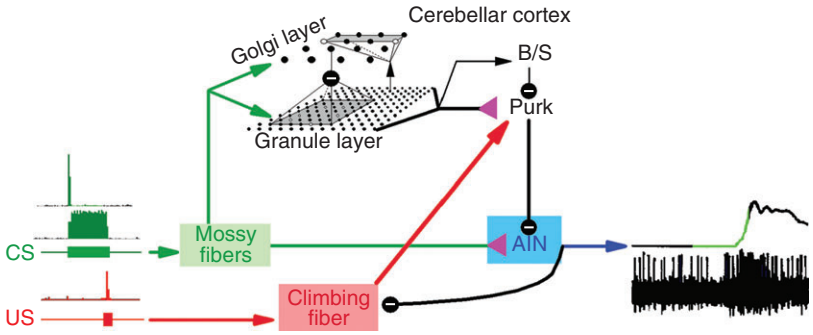


FIGURE 31.9 Circuit underlying eyelid conditioning. The conditioned (CS; tone) and unconditioned stimuli (US; air puff) activate mossy and climbing fibers, respectively. Some mossy fibers fire throughout the duration of the tone, while others fire only at the onset (green). Mossy fibers diverge to granule and Golgi cells in the granule layer, and to neurons in the anterior interpositus nucleus (AIN). The granule cells influence basket/stellate (B/S) and Purkinje cells, which in turn modulate AIN firing. Plasticity at the granule cell to Purkinje synapse, as well as the mossy fiber to nucleus synapse, appears to occur during learning. As a result, AIN neurons are thought to see a powerful barrage of Purkinje inhibition during the beginning of the tone. The relief of inhibition combined with mossy fiber excitation at the appropriate time triggers AIN firing, which (via the red nucleus; not shown) causes motor neuron firing to close the eyelid.

Feedforward Prediction Contributes to Cerebellar Coordination, Timing, and Learning

Analyses of eyelid conditioning and other behaviors suggest a common computational framework that ties together many of the functions proposed for the cerebellum. In all behaviors, sensory information can be used to coordinate, guide, and adjust movements. But most sensory systems are too slow to provide signals that can be used effectively to coordinate and control fast movements. The solution provided by the cerebellum is elegant: instead of using sensory signals to adjust ongoing movements, the cerebellum learns how to optimize subsequent movements. The computation performed by the cerebellum is thus a form of feedforward prediction.

In principle there are two ways to use sensory input to guide appropriate output: feedback and feedforward. Feedback is familiar and simple—for example, a thermostat. The measured temperature (sensory input) is compared to a target, and differences are translated into turning on the heater or cooler. Feedback systems of this sort can be accurate and reliable, but they cannot be fast. Attempts to make a feedback system fast result in

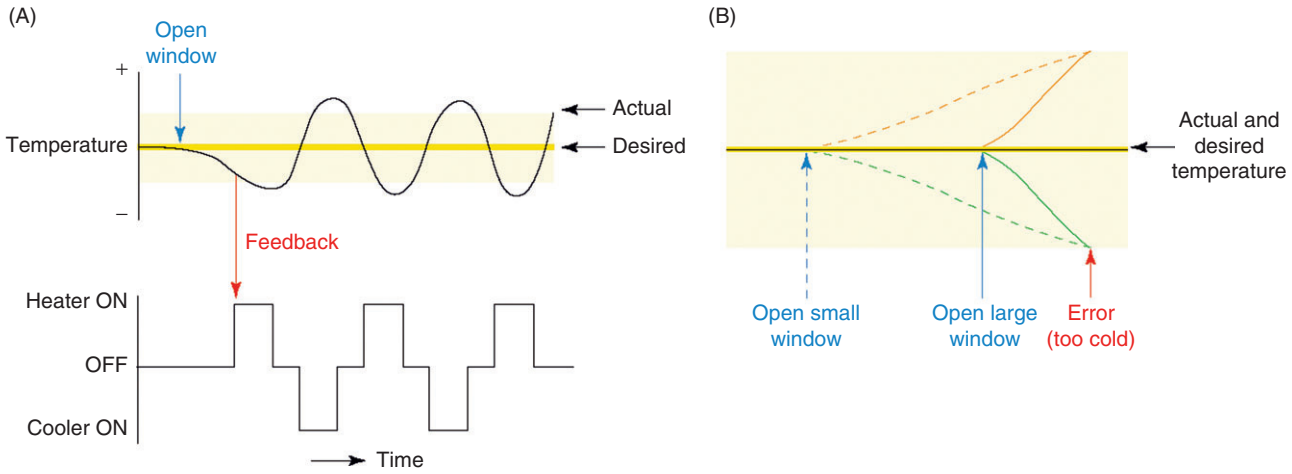


FIGURE 31.10 Feedback and feedforward control. (A) An example of a feedback-regulated thermostat. Though simple and accurate, feedback regulation tends to be slow, and attempts to speed it up cause oscillations. In this example, an attempted correction by a powerful heating system (in response to a decrease in temperature caused by opening a window) overshoots the mark, forcing another attempted adjustment by a powerful cooling system that also overshoots the mark, and so on. The upper and lower boundaries of the yellow area indicate the thresholds for detecting temperature change. (B) An example of a feedforward thermostat, which counteracts (orange lines) the changes in temperature predicted by the opening of windows (green lines). Though much quicker than feedback control, feedforward control is more complicated, requiring the ability to detect error-predicting stimuli (windows opening) and to delay anticipatory responses appropriately. Here, the opening of a large window (unbroken green line) predicts a faster decrease in temperature than does the opening of a small window (broken green line). The anticipatory response (increasing room temperature; broken or unbroken orange lines) must therefore be appropriately delayed in response to the predictive stimulus (opening a large or small window). This type of calibration requires associative learning that is temporally specific. *Modified from Ohyama et al. (2003).*

oscillations as, in the thermostat example, small deviations from the target produce alternating activations of the heater and cooler as each output causes overshoot of the target temperature (Fig. 31.10). Feedforward control can be faster because it involves anticipation of errors from previous experience. A hypothetical feedforward thermostat would be outfitted with a variety of sensory devices for temperature, humidity, number of people in the room, position of the sun, number of windows open, and so on. When a rapid change in room temperature is required, this hypothetical feedforward thermostat would predict, from previous experience, the burst of hot or cold air required to produce the rapid change.

If the thermostat's response yields the wrong outcome, then a feedforward thermostat should learn from this mistake so that subsequent predictions in similar circumstances are better. The properties required of this learning, as instantiated in the cerebellum, are fairly specific. Learning should produce changes in cerebellar signals that are timed to occur just before the error occurred (since that was the errant output). Cerebellar learning is thus expected to be associative: it should only modify the future output of the system for inputs that are similar to the ones that were present just before the error. In addition, learning should show temporal flexibility. The learned changes in output should not be time locked to the onset of any given mossy fiber input, but should be delayed after their onset so that the changed responses fix the errant prediction—namely, the one that occurred just before the arrival of the error signal. In other words, the error signal should correct the output that just occurred.

Eyelid conditioning studies reveal precisely these properties in cerebellar learning (Ohyama, Nores, Murphy, & Mauk, 2003). It is associative in that it takes mossy fiber inputs predicting climbing fiber inputs to occur. Moreover, the timing of the responses is precisely as described above. Pairing a mossy fiber input with a climbing fiber input produces a change in cerebellar output that is not time locked to the onset of the mossy fiber input but is instead timed to peak just before the input from the climbing fiber. This shows that cerebellar learning can develop predictions, through trial and error learning with climbing fiber error signals, that improve the prediction about the proper cerebellar output for a given mossy fiber input.

These considerations may tie together many key ideas about cerebellar function. Motor coordination is improved by feedforward predictions, which must be learned. This learning requires associative learning with quite specific temporal properties. It requires motor learning with timing. An important challenge for the future will be to determine how feedforward prediction contributes to the various motor, cognitive, and autonomic behaviors that engage the cerebellum.

Summary

Ideas about cerebellar function have been driven by analyzing the effects of lesions or dysfunction of the

cerebellum, by considering the nature of cerebellar inputs and outputs, and by examining the properties of cells and synapses within the cerebellar circuit. Analyses of cerebellar function that incorporate behavioral analyses, neuronal recordings, and computational perspectives are providing a framework for unifying a wide range of experimental results. The cerebellum appears to optimize downstream circuits by learning to predict the consequences of intended actions.

References

Bagnall, M. W., Zingg, B., Sakatos, A., Moghadam, S. H., Zeilhofer, H. U., & du Lac, S. (2009). Glycinergic projection neurons of the cerebellum. *Journal of Neurosciences, 29,* 10104–10110.

Bant, J. S., & Raman, I. M. (2010). Control of transient, resurgent, and persistent current by open-channel block by Na channel beta4 in cultured cerebellar granule neurons. *Proceedings of the National Academy of Sciences of the United States of America, 107,* 12357–12362.

Bastian, A. J. (2011). Moving, sensing and learning with cerebellar damage. *Current Opinion in Neurobiology, 21,* 596–601.

Best, A. R., & Regehr, W. G. (2009). Inhibitory regulation of electrically coupled neurons in the inferior olive is mediated by asynchronous release of GABA. *Neuron, 62,* 555–565.

Brody, B. A., Kinney, H. C., Kloman, A. S., & Gilles, F. H. (1987). Sequence of central nervous system myelination in human infancy. I. An autopsy study of myelination. *Journal of Neuropathology and Experimental Neurology, 46,* 283–301.

Carey, M. R. (2011). Synaptic mechanisms of sensorimotor learning in the cerebellum. *Current Opinion in Neurobiology, 21,* 609–615.

Carter, B. C., & Bean, B. P. (2011). Incomplete inactivation and rapid recovery of voltage-dependent sodium channels during high frequency firing in cerebellar Purkinje neurons. *Journal of Neurophysiology, 105,* 860–871.

Coesmans, M., Weber, J. T., De Zeeuw, C. I., & Hansel, C. (2004). Bidirectional parallel fiber plasticity in the cerebellum under climbing fiber control. *Neuron, 44,* 691–700.

De Zeeuw, C. I., Hansel, C., Bian, F., Koekkoek, S. K., van Alphen, A. M., Linden, D. J., & Oberdick, J. (1998). Expression of a protein kinase C inhibitor in Purkinje cells blocks cerebellar LTD and adaptation of the vestibulo-ocular reflex. *Neuron, 20,* 495–508.

De Zeeuw, C. I., Simpson, J. I., Hoogenraad, C. C., Galjart, N., Koekkoek, S. K., & Ruigrok, T. J. (1998). Microcircuitry and function of the inferior olive. *Trends in Neurosciences, 21,* 391–400.

Dugue, G. P., Dumoulin, A., Triller, A., & Dieudonne, S. (2005). Target-dependent use of co-released inhibitory transmitters at central synapses. *Journal of Neuroscience, 25,* 6490–6498.

Garcia, K. S., Steele, P. M., & Mauk, M. D. (1999). Cerebellar cortex lesions prevent acquisition of conditioned eyelid responses. *Journal of Neuroscience, 19,* 10940–10947.

Geurts, F. J., De Schutter, E., & Dieudonne, S. (2003). Unraveling the cerebellar cortex: Cytology and cellular physiology of large-sized interneurons in the granular layer. *Cerebellum, 2,* 290–299.

Gittis, A. H., Moghadam, S. H., & du Lac, S. (2010). Mechanisms of sustained high firing rates in two classes of vestibular nucleus neurons: Differential contributions of resurgent Na, Kv3, and BK currents. *Journal of Neurophysiology, 104,* 1625–1634.

Gould, B. B., & Graybiel, A. M. (1976). Afferents to the cerebellar cortex in the cat: Evidence for an intrinsic pathway leading from the deep nuclei to the cortex. *Brain Research, 110,* 601–611.

Holmes, G. (1939). The cerebellum of man. The Hughlings Jackson memorial lecture. *Brain, 62,* 1–30.

Holstein, G. R., Friedrich, V. L., Jr., Kang, T., Kukielka, E., & Martinelli, G. P. (2011). Direct projections from the caudal vestibular nuclei to the ventrolateral medulla in the rat. *Neuroscience, 175,* 104–117.

Hull, C., & Regehr, W. G. (2012). Identification of an inhibitory circuit that regulates cerebellar Golgi cell activity. *Neuron, 73,* 149–158.

Ito, M. (1996). Cerebellar long-term depression. *Trends in Neurosciences, 19,* 11–12.

Ito, M. (2001). Cerebellar long-term depression: Characterization, signal transduction, and functional roles. *Physiological Reviews, 81,* 1143–1195.

Ito, M. (2009). Functional roles of neuropeptides in cerebellar circuits. *Neuroscience, 162,* 666–672.

Ito, M., Sakurai, M., & Tongroach, P. (1982). Climbing fibre induced depression of both mossy fibre responsiveness and glutamate sensitivity of cerebellar Purkinje cells. *Journal of Physiology, 324,* 113–134.

Jorntell, H., & Hansel, C. (2006). Synaptic memories upside down: Bidirectional plasticity at cerebellar parallel fiber-Purkinje cell synapses. *Neuron, 52,* 227–238.

Khaliq, Z. M., & Raman, I. M. (2005). Axonal propagation of simple and complex spikes in cerebellar Purkinje neurons. *Journal of Neuroscience, 25,* 454–463.

Kolkman, K. E., McElvain, L. E., & du Lac, S. (2011). Diverse precerebellar neurons share similar intrinsic excitability. *Journal of Neuroscience, 31,* 16675–16674.

Lang, E. J., Sugihara, I., & Llinas, R. (1996). GABAergic modulation of complex spike activity by the cerebellar nucleoolivary pathway in rat. *Journal of Neurophysiology, 76,* 255–275.

Linden, D. J., Dickinson, M. H., Smeyne, M., & Connor, J. A. (1991). A long-term depression of AMPA currents in cultured cerebellar Purkinje neurons. *Neuron, 7,* 81–89.

Marr, D. (1969). A theory of cerebellar cortex. *Journal of Physiology, 202,* 437–470.

Mathy, A., Ho, S. S., Davie, J. T., Duguid, I. C., Clark, B. A., & Hausser, M. (2009). Encoding of oscillations by axonal bursts in inferior olive neurons. *Neuron, 62,* 388–399.

Mauk, M. D., & Donegan, N. H. (1997). A model of Pavlovian eyelid conditioning based on the synaptic organization of the cerebellum. *Learning & Memory, 4,* 130–158.

Medina, J. F., Nores, W. L., & Mauk, M. D. (2002). Inhibition of climbing fibres is a signal for the extinction of conditioned eyelid responses. *Nature, 416,* 330–333.

McElvain, L. E., Bagnall, M. W., Sakatos, A., & du Lac, S. (2010). Bidirectional plasticity gated by hyperpolarization controls the gain of postsynaptic firing responses at central vestibular nerve synapses. *Neuron, 68,* 763–775.

Miles, F. A., & Lisberger, S. G. (1981). Plasticity in the vestibulo-ocular reflex: A new hypothesis. *Annual Review of Neuroscience, 4,* 273–299.

Mugnaini, E., Sekerková, G., & Martina, M. (2011). The unipolar brush cell: A remarkable neuron finally receiving deserved attention. *Brain Research Reviews, 66,* 220–245.

Nishiyama, H., Fukaya, M., Watanabe, M., & Linden, D. J. (2007). Axonal motility and its modulation by activity are branch-type specific in the intact adult cerebellum. *Neuron, 56,* 472–487.

Ohyama, T., Nores, W. L., Murphy, M., & Mauk, M. D. (2003). What the cerebellum computes. *Trends in Neurosciences, 26,* 222–227.

Palkovits, M., Mezey, E., Hamori, J., & Szentagothai, J. (1977). Quantitative histological analysis of the cerebellar nuclei in the cat. I. Numerical data on cells and on synapses. *Experimental Brain Research, 28,* 189–209.

Person, A. L., & Raman, I. M. (2012). Purkinje neuron synchrony elicits time-locked spiking in the cerebellar nuclei. *Nature, 481,* 502–505.

Pugh, J. R., & Raman, I. M. (2009). Nothing can be coincidence: Synaptic inhibition and plasticity in the cerebellar nuclei. *Trends in Neurosciences, 32,* 170–177.

Raman, I. M., & Bean, B. P. (1999). Ionic currents underlying spontaneous action potentials in isolated cerebellar Purkinje neurons. *Journal of Neuroscience, 19,* 1663–1674.

Ray, R. S., & Dymecki, S. M. (2009). Rautenlippe redux—toward a unified view of the precerebellar rhombic lip. *Current Opinion in Cell Biology, 21,* 741–747.

Raymond, J. L., & Lisberger, S. G. (1996). Behavioral analysis of signals that guide learned changes in the amplitude and dynamics of the vestibulo-ocular reflex. *Journal of Neuroscience, 16,* 7791–7802.

Saviane, C., & Silver, R. A. (2006). Fast vesicle reloading and a large pool sustain high bandwidth transmission at a central synapse. *Nature, 439,* 983–987.

Schonewille, M., Gao, Z., Boele, H. J., Veloz, M. F., Amerika, W. E., Simek, A. A., De Zeeuw, C. I., et al. (2011). Reevaluating the role of cerebellar LTD in motor learning. *Neuron, 70,* 43–50.

Shadmehr, R., & Krakauer, J. W. (2008). A computational neuroanatomy for motor control. *Experimental Brain Research, 185,* 359–381.

Shadmehr, R., Smith, M. A., & Krakauer, J. W. (2010). Error correction, sensory prediction, and adaptation in motor control. *Annual Review Neuroscience, 33,* 89–108.

Shin, M., Moghadam, S. H., Sekirnjak, C., Bagnall, M. W., Kolkman, K. E., Jacobs, R., du Lac, S., et al. (2011). Multiple types of cerebellar target neurons and their circuitry in the vestibulo-ocular reflex. *Journal of Neurosciences, 31,* 10776–10786.

Spencer, R. M., Zelaznik, H. N., Diedrichsen, J., & Ivry, R. B. (2003). Disrupted timing of discontinuous but not continuous movements by cerebellar lesions. *Science, 300,* 1437–1439.

Steriade, M. (1995). Two channels in the cerebello-thalamo-cortical system. *Journal of Comparative Neurology, 354,* 57–70.

Strick, P. L., Dum, R. P., & Fiez, J. A. (2009). The cerebellum and nonmotor function. *Annual Review Neuroscience, 32,* 413–434.

Szapiro, G., & Barbour, B. (2009). Parasynaptic signalling by fast neurotransmitters: The cerebellar cortex. *Neuroscience, 162,* 644–655.

Telgkamp, P., Padgett, D. E., Ledoux, V. A., Woolley, C. S., & Raman, I. M. (2004). Maintenance of high-frequency transmission at purkinje to cerebellar nuclear synapses by spillover from boutons with multiple release sites. *Neuron, 41,* 113–126.

Urbano, F. J., Simpson, J. I., & Llinas, R. R. (2006). Somatomotor and oculomotor inferior olivary neurons have distinct electrophysiological phenotypes. *Proceedings of the National Academy of Sciences of the United States of America, 103,* 16550–16555.

Uusisaari, M., & Knopfel, T. (2010). GlyT2+ neurons in the lateral cerebellar nucleus. *Cerebellum, 9,* 42–55.

Wadiche, J. I., & Jahr, C. E. (2001). Patterned expression of Purkinje cell glutamate transporters controls synaptic plasticity. *Nature Neuroscience, 8,* 1329–1334.

Wang, V. Y., & Zoghbi, H. Y. (2001). Genetic regulation of cerebellar development. *Nature Reviews Neuroscience, 2,* 484–491.

Wang, Y. T., & Linden, D. J. (2000). Expression of cerebellar long-term depression requires postsynaptic clathrin-mediated endocytosis. *Neuron, 25,* 635–647.

Zhang, W., & Linden, D. J. (2003). The other side of the engram: Experience-driven changes in neuronal intrinsic excitability. *Nature Reviews Neuroscience, 4,* 885–900.

Martha Bagnall, Sascha du Lac, and Michael Mauk

CHAPTER

32

Eye Movements

As we learned in Chapter 26, the photoreceptor mosaic of the vertebrate retina transduces light energy in the form of photons into neural activity, ultimately in the form of action potentials. The spatial resolution of this transduction system is limited by the mosaic of photoreceptors and the optics of the eye, but only if the eye is held stationary with respect to the objects of interest in the external world. Thus, stabilizing the eyes with regard to the outside world and aiming the eyes toward moving or stationary targets are critical challenges to effective vision. Evolutionary pressures have shaped the eye movement systems of animals to meet this challenge in ways that are tailored to the visual structures and behavioral needs of each species. In this chapter we examine the behavioral and neural systems used by vertebrates to control eye movements in the support of vision.

As we shall see, the control of eye movements provides a window into several fundamental aspects of neuroscience. Because the eyes are a relatively simple mechanical system, eye movements provide an excellent system for investigating the neural mechanisms of motor control, and also offer some of the clearest views of other complex processes, such as sensory-motor interactions, and the higher-order control of behavior.

EYE MOVEMENTS ARE USED TO STABILIZE GAZE OR TO REDIRECT GAZE

Similar to the handling of a camera, the images that fall on the retina depend on how the eyes are held and moved. The orientation of the eyes in the head and the orientation and position of the head in space together determine the gaze direction (i.e., gaze = eye + head) and, consequently, control the retinal image. These spatial arrangements underlie the two general classes of eye movements in vertebrates. One class is responsible for *stabilizing gaze*. When animals move with respect to their surroundings, they risk degrading their visual acuity, because moving the head necessarily moves the eyes, and can cause images to streak across the retina. Gaze stabilization mechanisms have evolved to solve this problem. They maintain visual acuity during self-motion by stabilizing the retinal image of the world with rotations of the eyes that exactly compensate for head and body movements. The neural mechanisms for gaze stabilization are highly conserved across vertebrates, reflecting the widespread need to stabilize visual inputs despite other sensory and motor differences between species. The second class is responsible for *redirecting gaze*. When animals visually inspect their surroundings, they may use these eye movements to aim the line of sight at objects or features of particular interest, even when those items move. In contrast to stabilizing gaze, the process of redirecting gaze involves selectively sampling the animal's visual environment, because tracking one visual object often causes the retinal images of other objects to become blurred. Accordingly, the mechanisms for redirecting gaze are found only in animals that have retinal specializations, such as the primate fovea, that can be used to examine a limited region of visual space at a higher spatial resolution.

Stabilizing Gaze

There are two types of mechanisms for stabilizing gaze (Fig. 32.1): the *vestibulo-ocular reflex* (VOR) and the *optokinetic response* (OKR). The VOR uses signals from the vestibular labyrinth to counterrotate the eyeballs to keep retinal images stable during head movements. The *rotational VOR* is driven by head-rotation signals from the semicircular canals and represents a phylogenetically old reflex present in all vertebrate species. The fast transduction properties of the vestibular

© 2013 Elsevier Inc. All rights reserved.

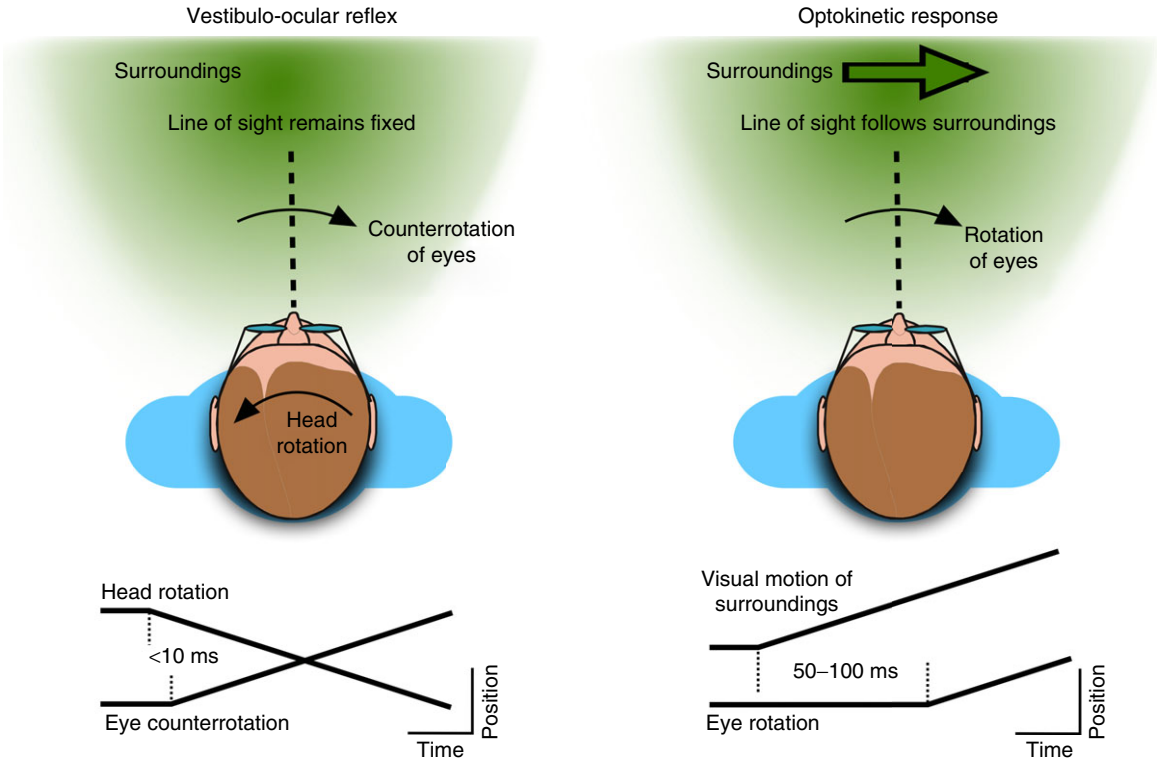


FIGURE 32.1 Eye movements that stabilize gaze. The vestibulo-ocular reflex keeps the line of sight fixed in the world by counterrotating the eyes during movements of the head. Here, the eyes rotate rightward at a short latency after the beginning of the leftward head movement. The optokinetic response stabilizes the line of sight with respect to the moving visual surround, but does so after a longer latency.

periphery, combined with a direct 3-neuron pathway, give the rotational VOR a very short latency (<10 ms). However, because the canals do not accurately transduce constant velocity head rotations, the output of the VOR decays over several seconds. Some vertebrates also possess a *translational VOR*, which is driven by head-movement signals from the otoliths. Unlike the image motion caused by head rotations, which is uniform across the visual field and can be fully compensated for by counterrotating the eyes, the image motion caused by translations are complex and can be only partly eliminated by eye movements. This explains why the translational VOR is found only in animals with central retinal specializations (e.g., foveal vision).

The *optokinetic response* (*OKR*) is driven by visual signals about the en masse movement of images across the retina. Because of the slow transduction of visual signals, the OKR has a longer latency (50–100 ms) than the VOR, but the use of visual motion signals makes it able to compensate for the constant motions and low speeds that the VOR cannot. The properties of the OKR therefore complement those of the VOR in maintaining stable gaze during movements, and the visual signals used by the OKR play a major role in guiding adaptive changes in the VOR. Animals with foveal vision also possess an *ocular following response* that is driven at very short latencies (~50 ms) by shifts in the retinal image, and that serves as the visual complement to the translational VOR (Miles, 1997).

Redirecting Gaze

Animals with foveal vision also possess several mechanisms for redirecting gaze (Fig. 32.2). *Saccadic* eye movements act to quickly move the image of a visual target from an eccentric retinal location to the center of the retina where it can be seen best. Saccades are very fast movements, reaching peak speeds greater than 500 deg/s and lasting only tens of milliseconds. Consequently, although saccades are often triggered by visual stimuli (at typical latencies of 150–250 ms), the trajectory of saccades is not guided by visual feedback, but instead follows a stereotyped profile that is largely determined by the size of the movement. Saccades do not require a visual stimulus, but can be guided by other modalities, and can even be directed toward the location of imagined or remembered targets.

Smooth pursuit eye movements slowly rotate the eyes to compensate for any motion of the visual target and thus act to minimize the blurring of the target's retinal image that would otherwise occur. Smooth pursuit is a relatively slow movement whose trajectory is determined by the moving stimulus. Pursuit is triggered by visual stimuli at latencies similar to, but somewhat shorter than, the latencies

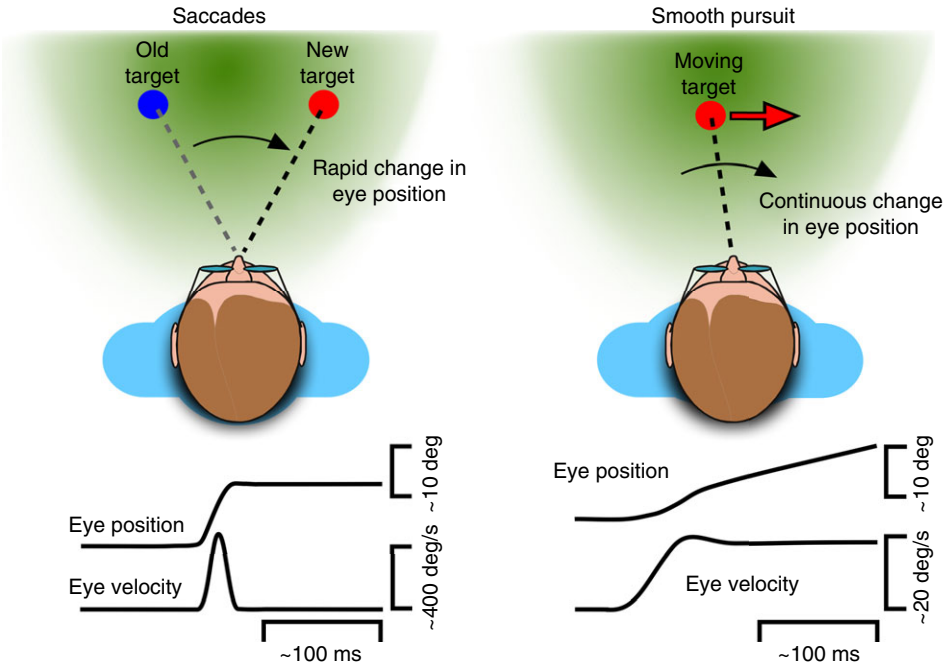


FIGURE 32.2 Eye movements that redirect gaze. Saccades change the line of sight to place the retinal image of visual targets onto the fovea. They are characterized by rapid changes in eye position (upward deflection in eye position trace) involving very high eye velocities (brief pulse in eye velocity trace). Smooth pursuit continuously changes the line of sight to minimize blurring of the target's retinal image. These movements are characterized by smooth and continuous changes in eye position (ramp in eye position trace) involving lower eye velocities (smooth step in eye velocity trace).

of saccades (typically 100–200 ms), and generates eye velocities up to ~50 deg/s, continuously adjusted by visual feedback about the target's retinal image. Unlike saccades, pursuit cannot be generated in the absence of a stimulus, although moving targets that are sensed through modalities other than vision can also guide pursuit.

Saccades and smooth pursuit produce conjugate, or yoked, movements of the two eyes. These are also referred to as *versional* eye movements. In contrast, a third type of mechanism for redirecting gaze, called *vergence* eye movements, produces disconjugate movements; the two eyes move in opposite directions or by different amounts. Vergence eye movements act to change the depth at which the eyes' lines of sight meet. When we look at a near object our eyes *converge* (rotate inward), and when we look at a far object, our eyes *diverge* (rotate outward). Before redirecting our gaze to an object at a different distance, the pair of images produced by that object will be spatially mismatched between the two eyes, causing binocular disparity. Also, because the curvatures of the two eyes' lenses are set to focus the currently viewed object, the image of the new object will also be blurred. These two signals, binocular disparity and blur, drive the combination of vergence eye movements and *accommodation* (which changes the dioptric power of the lenses by contracting or relaxing the ciliary muscle) that bring the new object into focus on the fovea of both eyes. Vergence and accommodation are tightly linked and are often referred to together as the *near response*. When evoked by themselves (e.g., by binocular disparity), vergence movements are very slow, taking up to 1 second to converge or diverge the eyes. However, when they occur in combination with faster versional eye movements such as saccades, vergence movements speed up, and interestingly, so does accommodation.

Fixation and Fixational Eye Movements

In between gaze movements, the eyes are actively held steady by a *fixation* mechanism. By preventing the eyes from redirecting gaze to another visual target, fixation makes it possible to visually inspect a particular object at length. Periods of fixation typically last about 200 ms but vary depending on what the animal is doing. Fixation does not necessarily act to keep the eyes from moving, but to hold the eyes steady with respect to the environment. For example, gaze stabilization mechanisms such as the VOR are active during periods of fixation.

Even during periods of steady fixation of a stationary object, the eyes are not completely still. The eyes typically make tiny movements during fixation in seemingly random directions that cause very slight changes in how images fall on the retinal mosaic but without displacing images from the fovea. *Microsaccades* are very small saccades with amplitudes typically less than 0.1 degree of visual angle, and are interspersed with *slow drifts* (speeds less than ¼ degree per second) of similar amplitude. The function of these movements remains unclear. However, like other larger saccades, they can prevent the adaptation of retinal signals that would occur with perfectly stabilized images, they are influenced by shifts of attention, and they can be strategically targeted or suppressed depending on the task.

BOX 32.1

EYE MOVEMENTS AND READING

During reading, saccadic eye movements bring successive snippets of text images to the center of the retina. The amplitude of the saccades depends on the letter size—a typical reader makes saccades that shift the line of sight by about 8 character spaces. In common situations—for example, viewing 12-point font at a distance of about 18 inches—each saccade is therefore only a couple of degrees in amplitude. The majority of saccades (~85%) form a forward path along the line of text (e.g., left to right for English readers), but a substantial minority of saccades (~15%), referred to as *regressions*, direct the line of sight back to text that has already been foveated.

The periods of fixation between saccades are crucial for reading comprehension, and typically last only about one-fifth of a second (~200 ms). Fixation provides the high-acuity input about individual letters—for example, telling the difference between "o" and "c"—that is necessary to correctly identify words. The number of letters that can be processed during each fixation (the *perceptual span*) is a key factor in determining reading proficiency. In skilled readers, the perceptual span stretches ahead of the currently fixated point, covering about 14 letters forward and only about 4 letters backward. Letters at the center of the perceptual span are processed for meaning and identifying words; letters at the edges of the perceptual span appear to be important for parsing the text and perhaps determining the landing points for upcoming saccades.

The skillful deployment of saccades and fixation for reading requires a great deal of practice, and represents an important developmental milestone for visual-motor control and cognition. In typical individuals, as text becomes harder to read, the amplitudes of saccades decrease, the number of regressions increase, and the duration of fixations increase. In reading disorders, such as dyslexia, these patterns are exaggerated, and accompanied by other unusual scan path features, such as targeting the beginning rather than the middle of words. This situation poses a chicken-or-the-egg problem: are reading disorders caused by a problem with controlling eye movements, or are eye movement problems caused by the reading disorder?

THE MECHANICS OF MOVING THE EYES

Each Eye Is Controlled by Three Pairs of Muscles

Moving the eye involves rotating the globe of the eye in the orbital socket of the skull. Because the eye undergoes minimal translation during movement, it can be viewed as a spherical joint with an orientation defined by three axes of rotation (horizontal, vertical, and torsional). Correspondingly, the orientation and movement of each eye is accomplished by three complementary pairs of extraocular muscles (Fig. 32.3). The lateral and medial rectus muscles form an antagonistic pair that controls the horizontal position of the eye. Contraction of the lateral rectus (and commensurate relaxation of the medial rectus) causes the eye to abduct, or rotate outward. Conversely, contraction of the medial rectus (and relaxation of the lateral rectus) causes the eye to adduct, or rotate inward. The actions of the other two pairs of muscles are more complex. When the eye is centered in the orbit, the primary effect of the superior and inferior recti is to elevate (rotate up) or depress (rotate down) the eye. However, when the eye is deviated horizontally in the orbit, these muscles also contribute to torsion, the rotation of the eye around the line of site that determines the orientation of images on the retina. The primary effect of the superior and inferior obliques is intortion (the top of the eye rotates toward the nose) or extortion (the top of the eye rotates away from the nose), but depending on the horizontal deviation of the eye, these muscles may also help determine the vertical orientation of the eye. Eye movements therefore involve redefining the 3-dimensional orientation of the eye by changing the balance of forces applied to the globe by all three pairs of extraocular muscles.

The mechanics of moving the eye may seem simple, but actually provide a computationally complex challenge for the brain. Because the eye is free to rotate in 3-dimensions, the final orientation of the eye depends on the order in which the rotations are carried out—for example, "turn-then-spin" gives a different outcome than "spin-then-turn." This property is referred to as *noncommutativity*. In fact, when the eye moves, the final 3-dimensional orientation of the eye and the axis of rotation used to get there are not arbitrary, but form an orderly set. *Donder's law* states that the orientation of the eye is always the same when the eye is aimed in a particular direction. *Listing's law* specifies these orientations by stating that the axes used to rotate the eye are confined to a single plane, referred to as Listing's plane. These orderly relationships are achieved, at least in part, by a mechanism that involves the muscles themselves (Klier, Meng, & Angelaki, 2011). The extraocular muscles consist of two distinct layers: *global* fibers, which attach directly to the globe of the eye, and *orbital* fibers,

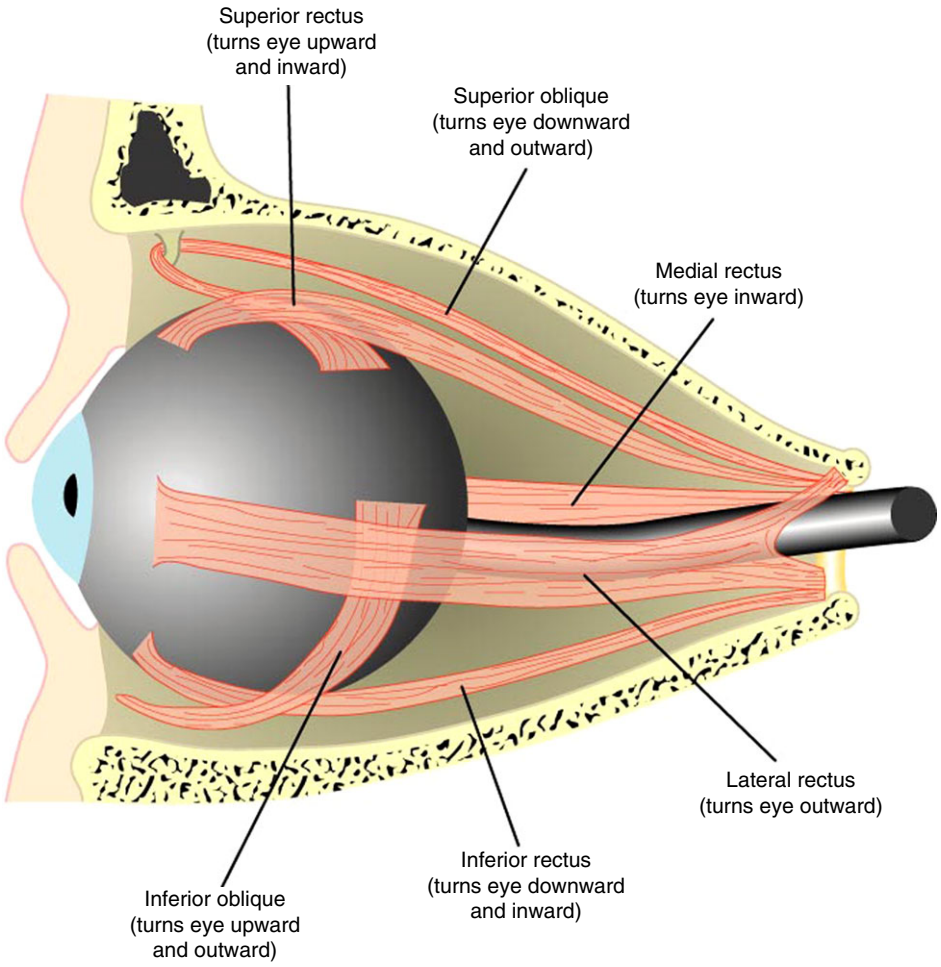


FIGURE 32.3 Muscles of the eye. Eye movements are controlled by six extraocular muscles arranged in three pairs, shown here in a cutaway view of the eye in its socket, or orbit.

which terminate in connective tissue that ensheathes the body of the muscle and acts like a moveable pulley for the global fibers. Because the pulleys themselves move when the muscle is contracted or relaxed, the pulling direction of the extraocular muscles changes with eye position. These gaze-dependent mechanical changes in muscle action largely account for Listing's law (Demer, 2006).

Interestingly, Listing's law is obeyed during saccades and smooth pursuit, but is violated during the VOR and the OKR. This difference presumably stems from the different inputs for these movements. Saccades and smooth pursuit may follow Listing's law because they involve local 2-dimensional retinal inputs that do not uniquely specify the orientation of the eye. In contrast, the VOR is driven by 3-dimensional inputs from the semicircular canals for which the extra degree of freedom is not redundant. Thus, the mechanics of the eye muscles impose kinematic constraints that are helpful during visually guided eye movements, but these constraints can be altered or lifted by neural control signals during gaze stabilization.

Central Control of the Extraocular Muscles

The six extraocular muscles are innervated by three of the bilaterally paired cranial nerves (Fig. 32.4). The oculomotor nerve (cranial nerve III) innervates the medial rectus, the superior and inferior rectus, and the inferior oblique on one side of the head. The cell bodies for these motor neurons thus lie in the third cranial nerve nucleus. The oculomotor nerve also contains the fibers that control accommodation of the lens. The trochlear nerve (cranial nerve IV) innervates the superior oblique muscle, and the abducens nerve (cranial nerve VI) innervates the lateral rectus. These three pairs of nuclei, distributed through the brain stem, contain all of the oculomotor motor neurons and are heavily interconnected by a pathway called the medial longitudinal fasciculus (mlf). This interconnection is crucial for the precise coordination of the extraocular muscles that is necessary for the control of eye movements.

The firing rates of the oculomotor motor neurons determine the forces applied by the extraocular muscles and, consequently, whether the eye moves or is held in place (Fuchs & Luschei, 1970). When the eye is stationary, the

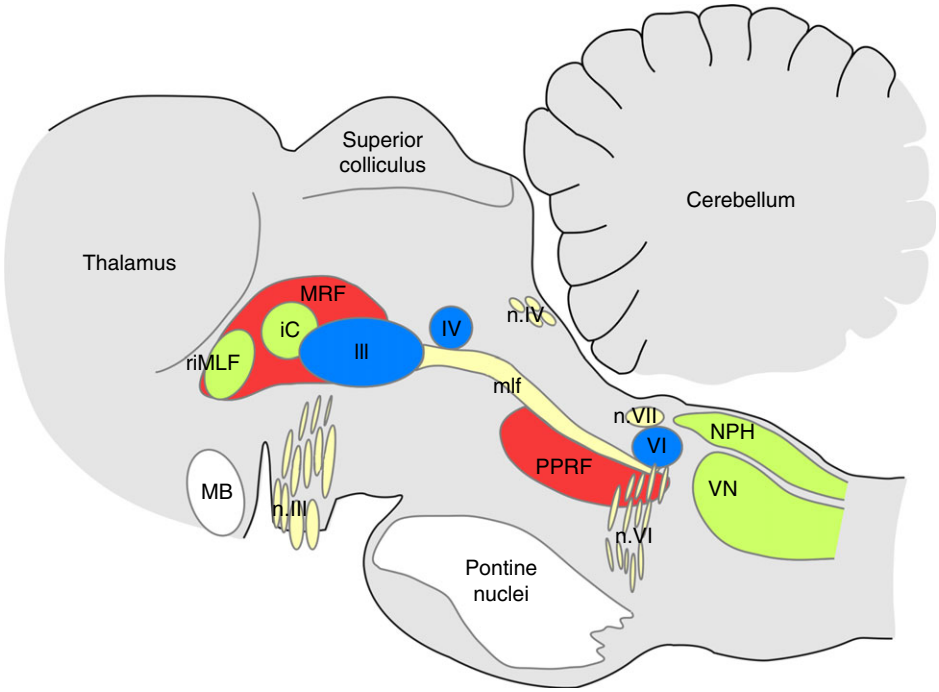


FIGURE 32.4 Oculomotor nuclei in the brainstem. Parasagittal section through the brainstem, cerebellum, and thalamus of a rhesus monkey, showing the location of the major brainstem nuclei involved in the control of eye movements. Motor neurons for the eye muscles are located in the oculomotor nucleus (III), trochlear nucleus (IV), and abducens nucleus (VI), and reach the extraocular muscles via the corresponding nerves (n. III, n. IV, n. VI). Premotor neurons for controlling eye movements are located in the paramedian pontine reticular formation (PPRF), the mesencephalic reticular formation (MRF), rostral interstitial nucleus of the medial longitudinal fasciculus (riMLF), the interstitial nucleus of Cajal (IC), the vestibular nuclei (VN), and the nucleus prepositus hypoglossi (NPH). The medial longitudinal fasciculus (mlf) is a major fiber tract containing the axons of these neurons. Other abbreviation: MB, mammillary body of the hypothalamus.

firing rate of oculomotor motor neurons is proportional to eye position, because under static conditions the balance of forces exerted on the globe determines the orientation of the eye (Fig. 32.5). Each motor neuron has a recruitment point (an eye position at which the motor neuron begins to fire) and a characteristic slope (the change in firing rate as the eye rotates in the pulling direction of the relevant muscle). Thus, as in the skeletal system, oculomotor motor neurons have a fixed recruitment order. When the eye moves in the pulling direction of the motor unit, oculomotor motor neurons exhibit a high-frequency pulse of activity that drives the high-velocity phase of the eye movement. The high-frequency pulse is followed by a sustained step-like increase in firing rate appropriate for the new static position of the eye. When the eye moves in the opposite direction, the motor neurons pause their firing during the movement and then resume their activity at a lower sustained level.

The neural commands from these motor neurons therefore take the form of a pulse-step during rapid eye movements such as saccades. The amplitude and duration of the pulse determine the speed and duration of the

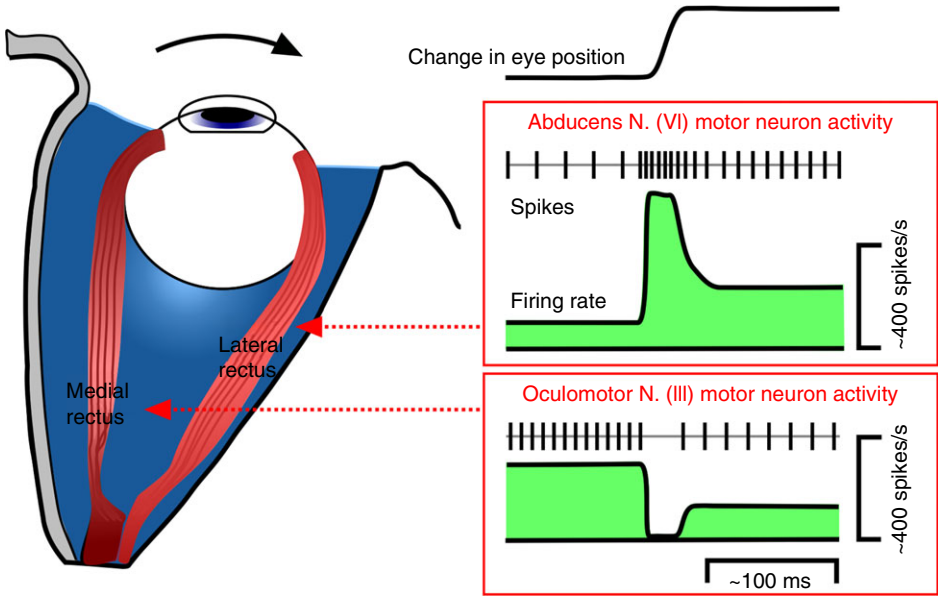


FIGURE 32.5 The firing rate patterns of eye movement motor neurons during horizontal eye movements. The rapid rightward change in eye position, indicated by the upward deflection in the eye position trace, involves contraction of the right lateral rectus and relaxation of the right medial rectus. This is accomplished by a pulse-step increase in activity of the agonist motor neurons in the abducens nucleus, and a pause-step decrease in activity of the antagonist motor neurons in the oculomotor nucleus.

movement, and the amplitude of the step determines the final static eye position. The need for the pulse derives from the sluggish dynamics of the eye and orbital tissues, which is dominated by the viscosity of the extraocular muscles. If the eye were moved by only a step change in force, it would take nearly a second for the eye to settle at its new orientation, which would obviously be incompatible with the goal of supporting vision. Similarly, during slow eye movements such as smooth pursuit, the neural commands from the motor neurons consist of a ramp-like change in activity to drive the smooth changes in eye position, plus an additional increment to overcome the sluggish dynamics of the eye. These patterns of activity are precisely adjusted by central mechanisms involving the brainstem and cerebellum, and illustrate a general problem in motor control—the need to convert desired actions into appropriately formed motor commands. The solution appears to involve the use of internal models that allow central mechanisms to take into account the dynamic properties of the body part or object to be moved.

Oculomotor motor units are also unusual in several ways (Buttner-Ennever, 2007). Unlike skeletal motor units, eye motor neurons participate equally in all types of eye movements, regardless of the speed involved. The extraocular muscles contain muscle spindles, but no Golgi tendon organs, and there are no ocular stretch reflexes. Instead, eye proprioception appears to mainly come from palisade endings on the eye muscle tendons. The function of this input is unclear, but it is distributed via the trigeminal nucleus to many key structures in the oculomotor system (including the vestibular nuclei, cerebellum, superior colliculus, and frontal eye fields) and may play a role in the long-term adaptive control of eye movements.

THE FUNDAMENTAL CIRCUITS FOR STABILIZING GAZE

The Vestibulo-Ocular Reflex (VOR) Stabilizes Gaze during Head Movements

The function of the VOR is to stabilize retinal images during head movements by using head-motion signals from the vestibular organs. A purely rotational VOR can be elicited in complete darkness by rotating the head roughly about its center (e.g., the interaural axis). Head rotations stimulate the semicircular canals of the inner ear (Chapter 29), which provide the sensory signals about head motion that drive this highly conserved mechanism for counterrotating the eyes. During sustained rotations of the head, the eyes cannot continue to counterrotate endlessly in the orbit, but are intermittently and rapidly reset to a central position during the ongoing compensatory eye rotation. The profile of eye position over time during the VOR therefore exhibits a characteristic sawtooth pattern consisting of the slow-phase counterrotations and quick-phase resetting eye movements. This pattern is called *nystagmus*, and is described by the direction of the quick phases (Fig. 32.6 is an example of right-beat nystagmus).

The speed of the slow phases during nystagmus indicates how effectively the VOR is compensating for the head rotation. The degree of compensation is often characterized by the VOR gain, which is defined as the ratio of slow-phase eye speed over the speed of head rotation. Perfect compensation occurs when the eye rotates in the opposite direction at exactly the same speed as the head rotation, which corresponds to a gain of 1. The ability of the VOR to compensate for head motion is limited by the dynamics of the signals from the semicircular canals, which act as "high-pass" sensors of head velocity. Thus, the VOR is very effective (its gain is near 1) for fast back-and-forth head rotations that take a second or less to complete (i.e., frequencies over 1 Hz), but it is much less effective for slow head rotations that continue over many seconds (e.g., 0.1 Hz). Under natural conditions, this limitation of the VOR does not pose a problem for vision, because at these lower speeds visual signals can be used to stabilize gaze very effectively through the optokinetic response.

The fundamental neural mechanism for the VOR consists of brainstem circuits linking the vestibular afferents to the extraocular muscles, including a direct 3-neuron arc pathway (Fig. 32.6). For example, during a rightward head rotation vestibular afferents from the horizontal semicircular canal on the right side increase their firing rates, and this increase is passed on to neurons in the vestibular nucleus (VN) on the same side. These vestibular nucleus neurons, in turn, increase the activity of motor neurons and interneurons in the abducens nucleus (VI) on the opposite (left) side. Motor neurons in the abducens innervate the lateral rectus of the left eye, completing the 3-neuron arc. Interneurons in the abducens provide a crossed projection back to the right oculomotor nucleus, where motor neurons innervate the medial rectus of the right eye. By contracting both the left lateral rectus and the right medial rectus, the two eyes rotate leftward together in response to the rightward head rotation. In addition, because the vestibular afferents on the left side decrease their activity during a rightward rotation, a complementary circuit causes relaxation of the right lateral rectus and the left medial rectus. The VOR is therefore controlled in a push-pull fashion that yokes the movements of the two eyes through the bilateral symmetry of these basic circuits.

These direct pathways are key components of the VOR mechanism, but the VOR also requires indirect pathways to function correctly. As discussed in the previous section, the firing rate of oculomotor motor neurons is primarily related to eye position, but the afferent signals from

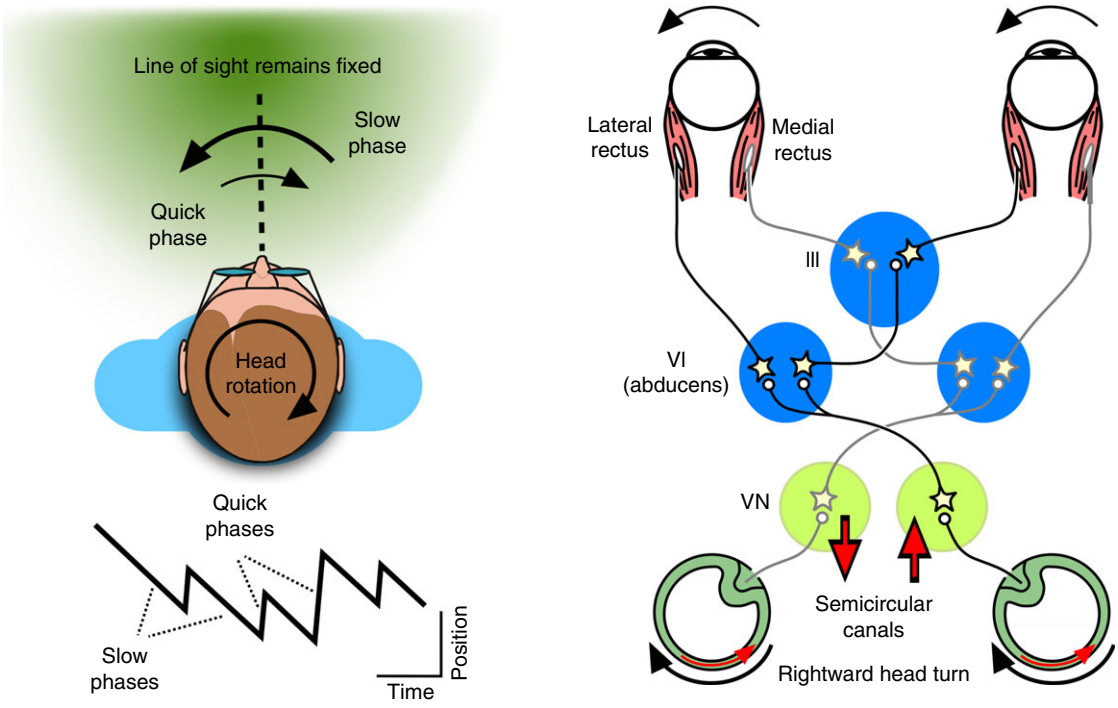


FIGURE 32.6 The eye movements and neural circuits for the VOR. During a sustained rotation of the head (here, rightward), the eyes rotate slowly to the left (downward deflections in eye position trace labeled "slow phases"), and then abruptly shift rightward to reset the eyes toward a central position (upward deflections in eye position trace labeled "quick phases"). This is an example of right-beat nystagmus. The diagram of the basic VOR circuit shows the push-pull organization of the horizontal VOR. When the head rotates rightward, activity increases in the right semicircular canal, right vestibular nucleus (VN), and left abducens nucleus (VI). This leads to increased activation of the left eye lateral rectus and, through a crossed projection to the oculomotor nucleus (III), increased activation of the right eye medial rectus. The complementary set of pathways (colored in gray) show decreases in activity.

the semicircular canals are primarily related to the head velocity. In order to construct the motor commands needed to hold, as well as move, the eyes during the VOR, the brain needs to transform the dynamic signals provided by the vestibular afferents into the combination of dynamic and static commands found on motor neurons. Similar to the operation of mathematical integration in calculus, this process involves the temporal integration of velocity inputs into position signals.

The *neural integrator* for eye movements is accomplished by neurons in the vestibular nuclei and two other brainstem nuclei—the nucleus prepositus hypoglossi (for horizontal components) and the interstitial nucleus of Cajal (for vertical)—and is fine-tuned by connections between these brainstem nuclei and the vestibulocerebellum (Cannon & Robinson, 1987). The output from the neural integrator pathway sums with the signals from the direct pathway to provide the motor neurons with the complete set of velocity and position commands needed to accomplish the VOR. Importantly, all conjugate eye movements also use this same indirect integrator pathway. Thus, although the source of the velocity signals for the VOR, OKR, saccades, and pursuit are quite different, they appear to share a common brainstem mechanism for transforming these signals into the final motor commands.

The Optokinetic Response (OKR) Uses Visual Inputs to Stabilize Gaze

The OKR also functions to stabilize retinal images during head movements, but uses visual inputs to infer the direction and speed of head motion rather than responding directly to head-velocity signals. The optimal stimulus for the OKR is full-field motion of the visual environment that, like head rotations, causes a saw-tooth pattern of eye movements referred to as *optokinetic nystagmus*. Like the VOR, the slow phase movements during nystagmus compensate for the head rotation that presumably caused the full-field motion. However, because the OKR is driven by visual signals, which are processed much more slowly than vestibular signals, the range of motions that are effectively compensated by the OKR is quite different from the VOR. The OKR is very effective (its gain is near 1) for slow rotations that continue over many seconds, but it is much less effective for very fast back-and-forth rotations. These properties of the OKR therefore complement those of the VOR and together, the two types of mechanisms can effectively stabilize gaze over a much wider range of speeds and behavioral conditions than either could achieve alone.

The interlocking nature of the two systems is also evident in the neural pathways for the OKR, despite the

difference in their input signals. In vertebrate species without foveal vision, such as the rabbit, the visual processing for the OKR is accomplished entirely within the brainstem. Direction-selective retinal ganglion cells project directly to a set of visual nuclei in the midbrain called the pretectum and accessory optic nuclei. In species with foveal vision, the visual properties of these nuclei are determined not only by inputs from the retina, but are also shaped by inputs from visual areas of the cerebral cortex. Neurons in the accessory optic nuclei have very large receptive fields and respond selectively to global visual motion that would result from rotating the head about a particular spatial axis. Remarkably, the selectivity of individual neurons is restricted to one of the three axes of rotation that are also detected by the semicircular canals. This common rotational coordinate system makes it easier to combine visual *reafferent* signals about head motion from the accessory optic nuclei with the vestibular *afferent* signals provided by the labyrinths. Accordingly, neurons in the pretectum and accessory optic nuclei project to the vestibular nuclei and nucleus prepositus hypoglossi, where these visual signals join the same final pathways used by the VOR. These nuclei also project to the vestibulocerebellum, where the visual signals contribute to the velocity commands for the OKR and other smooth eye movements, and also act as feedback signals that play a critical role in the long-term adjustment of the VOR.

THE COMMANDS FOR REDIRECTING GAZE ARE FORMED IN THE BRAINSTEM

The commands for redirecting gaze use the same final motor pathways that are used for stabilizing gaze, but the signals and mechanisms giving rise to the movement commands are quite different. For smooth pursuit, they are based on motion signals about the visual target that are first extracted in specialized areas of the cerebral cortex and then used to commandeer the brainstem and cerebellar pathways for the OKR. For saccades, which are not guided by visual feedback, a specialized circuit in the brainstem reticular formation produces the stereotyped velocity command that determines the trajectory of the saccade.

The Motor Commands for Saccades Are Constructed in the Brainstem Reticular Formation

The control signals for saccades descending from higher centers specify the location of the target, which is then transformed by circuits in the brainstem reticular formation into the motor commands for saccades (Fig. 32.7). The horizontal and vertical components of saccades are formed by separate but interconnected circuits—the horizontal commands are formed in the paramedian pontine reticular formation (PPRF), and the vertical commands are formed in a part of the mesencephalic reticular formation called the rostral interstitial nucleus of the medial longitudinal fasciculus (riMLF).

Within these brainstem circuits, often referred to as the *saccadic burst generator*, several different classes of neurons work together to construct the saccade motor command (Keller, 1974). *Burst neurons* fire a high-frequency volley of action potentials that begins ~10 ms before the onset of the saccade and ends slightly before the saccade lands on the target. The firing rate during the burst is related to the instantaneous velocity of the saccade, and the number of spikes in the burst is related to the amplitude of the saccade. There are several types of burst neurons. Excitatory burst neurons (EBN), sometimes called "medium-lead burst neurons," provide the velocity signal for the eye muscles that will contract during the saccade, and specify the "pulse" portion of the saccade motor command. The EBNs provide a copy of this velocity signal to the brainstem nuclei of the neural integrator, which then generates the "step" portion of the motor command, just like they do for other eye movements. EBNs also contact inhibitory burst neurons (IBN), which provide a crossed projection that inhibits motor neurons for the antagonist eye muscles that will relax during the saccade. Finally, long-lead burst neurons (LLBN) show changes in their activity well in advance of the saccade, presumably reflecting inputs received from higher centers, and provide an excitatory input to the EBNs.

The other major class of neurons in the saccadic burst generator is the *omnipause neurons* (OPN). These neurons fire at a constant high rate (~100 spikes/s) during fixation, but completely pause their activity just before and during saccades in all directions. Because the OPNs directly inhibit burst neurons, their tonic activity prevents unwanted saccades during fixation, and the pause in their activity unleashes the synchronized bursts of action potentials that initiate and drive the saccade. The onset of the pause in OPN activity appears to be triggered by inhibitory signals originating from higher brain centers, but the duration of the pause does not determine when the saccade ends. Instead, stopping saccades involves another inhibitory influence on burst neurons. One likely candidate is the oculomotor cerebellum (lobules VI and VII of the vermis and the caudal fastigial nucleus). The oculomotor vermis is interconnected with the saccadic burst generator and plays a crucial role in adjusting the metrics of saccades so that the eyes land accurately on the target at the end of the gaze movement.

The Motor Commands for Pursuit Are Formed in the Brainstem and Cerebellum

The control signals for pursuit descending from higher centers continuously specify the motion of the target to be followed with the eyes. Thus, in contrast to saccades,

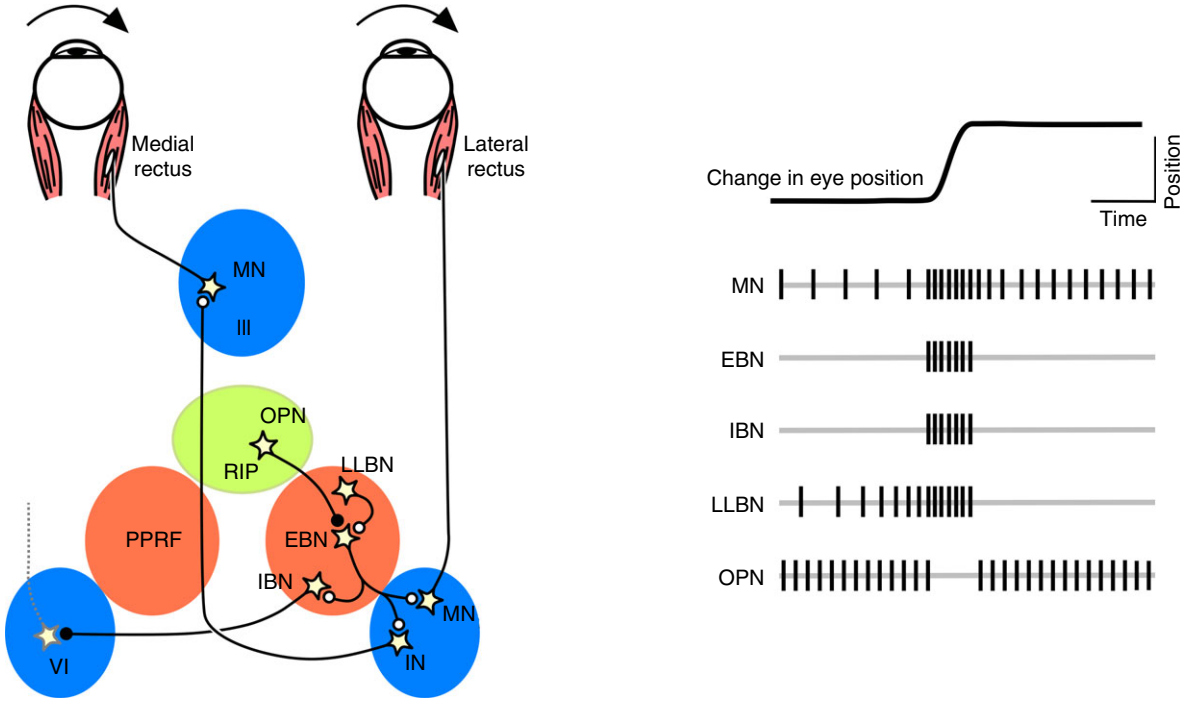


FIGURE 32.7 The brainstem circuits for controlling saccadic eye movements. This diagram of the saccade burst generator outlines the major classes of neurons involved in constructing the motor command for horizontal saccades. Omnipause neurons (OPN) located near the midline in the nucleus raphe interpositus (RIP) tonically inhibit excitatory burst neurons (EBN) located in the paramedian pontine reticular formation (PPRF). When OPNs pause, the EBNs emit a burst of spikes, which activate motor neurons (MN) in the abducens nucleus (VI) innervating the lateral rectus muscle. The burst also activates interneurons (IN) which, in turn, activate motor neurons on the oculomotor nucleus (III) on the opposite side, innervating the medial rectus. Inhibitory burst neurons (IBN) show a pattern of activity similar to EBNs, but provide inhibitory inputs to decrease activation in the complementary circuits and antagonist muscles. Long-lead burst neurons (LLBN) show activity long before movement onset and provide an excitatory input to EBNs.

pursuit does not require motor circuitry to construct the velocity command from scratch. Instead, the pursuit pathways in the brainstem and cerebellum sculpt the descending control signals to form an appropriately shaped velocity command, and provide a copy of this velocity signal to the neural integrator to generate the position component of the motor command.

The velocity commands for pursuit arise in areas of the cerebral cortex and reach the final motor circuits through several routes (Fig. 32.8). One pair of pathways involves projections from the cerebral cortex to the cerebellum via a large set of relay neurons on the floor of the brainstem called the basal pontine nuclei. Areas in extrastriate visual cortex provide signals related to visual motion through the dorsolateral pontine nuclei to a portion of the vestibulocerebellum called the ventral paraflocculus. The frontal cortex provides signals related to target and eye velocity through the dorsomedial pontine nuclei, and the adjacent nucleus reticularis tegmenti pontis, to the oculomotor cerebellum (vermis), the same portion of the cerebellum involved in adjusting the metrics of saccades. Both of these cerebellar regions provide inputs to parts of the brainstem motor pathways that we have already discussed. The output of the ventral paraflocculus is closely related to eye velocity and terminates in the vestibular nuclei, where it contributes to the velocity command for pursuit, and also provides an input to the neural integrator. The output of the oculomotor vermis is less well defined but appears to be related to eye acceleration as well as eye velocity. The outputs of the vermis and the associated deep cerebellar nucleus (the caudal fastigial nucleus) are sent to neurons in the brainstem reticular formation, and may play a role in shaping the velocity command appropriately when the target motion changes.

Another route for conveying motion signals for pursuit involves the pretectum. As described above, the pretectum plays a key role in the OKR, but in primates the pretectum receives inputs from areas of the cerebral cortex that are important for pursuit. Accordingly, some neurons in the primate pretectum have smaller receptive fields and respond well to the small moving stimuli typically used to elicit pursuit. The pretectum provides an input to the cerebellum through the basal pontine nuclei, but also projects directly to the vestibular nuclei and nucleus prepositus hypoglossi, offering a relatively direct route to the final motor pathways.

The brainstem control of pursuit also appears to involve some of the same neurons that are part of the saccadic

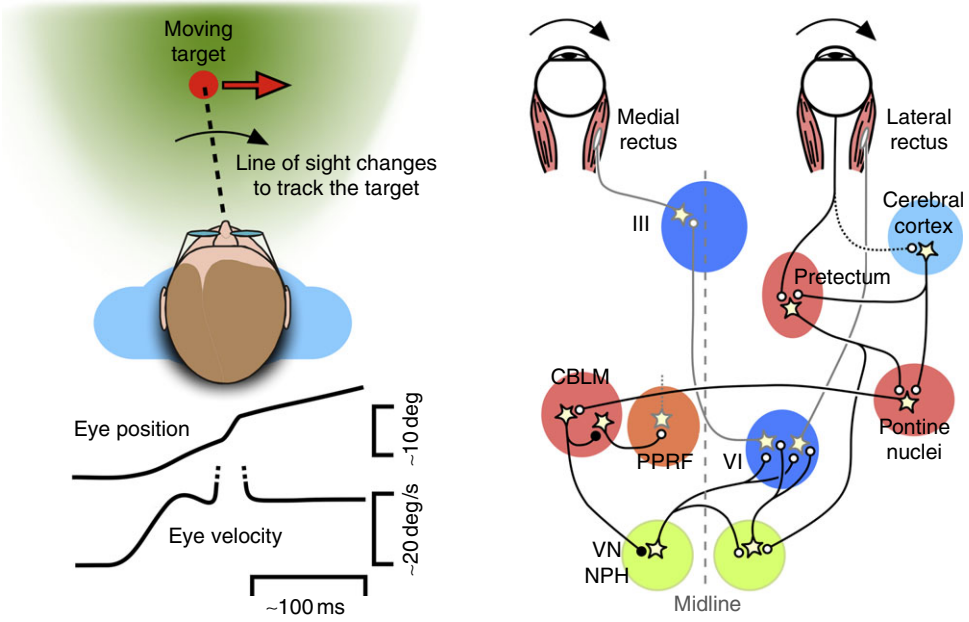


FIGURE 32.8 The brainstem and cerebellar circuits for controlling pursuit eye movements. During pursuit, the line of sight changes smoothly to follow the motion of the target. The smooth changes in eye velocity are often accompanied by small "corrective" saccades, indicated by the upward deflection in the eye velocity trace (dotted lines). The diagram outlines the circuits involved in constructing the motor command for horizontal pursuit. The major pathways involve projections from areas of the cerebral cortex via the pontine nuclei to the cerebellum (CBLM), which then modulates activity in nuclei associated with the VOR—namely, the vestibular nuclei (VN) and nucleus prepositus hypoglossi (NPH)—and also modulates activity in nuclei associated with saccades, such as the paramedian pontine reticular formation (PPRF). A second pathway involves projections from the pretectum to the pontine nuclei, as well as a direct projection to the vestibular nuclei and prepositus hypoglossi.

burst generator. The EBNs and IBNs are not involved in forming the pursuit motor command, but some of the long-lead burst neurons in the brainstem reticular formation increase their tonic firing during pursuit, and show a buildup of activity before the onset of pursuit and saccades. Even more surprisingly, OPNs—the "gatekeepers" for saccades—are also modulated during pursuit. They do not completely pause their activity as they do during saccades, but decrease their activity by about one-third. Moreover, electrical stimulation applied to the OPNs slows down pursuit eye movements, in addition to stopping saccades (Missal & Keller, 2002). The function of this overlap between saccades and pursuit in these premotor circuits is not yet fully understood, but one possibility is that descending information about the target does not directly trigger saccades and pursuit, but instead a motor decision process at the level of the brainstem and cerebellum determines which type or combination of movements is most appropriate under the particular circumstances.

GAZE MOVEMENTS ARE CONTROLLED BY THE MIDBRAIN AND FOREBRAIN

The motor commands for gaze movements formed in the brainstem depend on control signals descending from higher centers, in particular, from the superior colliculus (SC) and several eye-movement related areas of the cerebral cortex (Fig. 32.9). These brain regions make it possible to bring the well-honed motor circuits in the brainstem under the control of higher-order processes such as attention, perception, and cognition.

The Superior Colliculus Contains a Retinotopic Map for Controlling Gaze

The superior colliculus (SC) is a multilayered structure lying on the roof of the midbrain. In nonmammals, the homologous structure is called the optic tectum, which serves as the primary center for processing retinal inputs and for transforming sensory inputs into commands for orienting movements. In mammals, the SC has a similar functional role, and consists of superficial layers that receive direct inputs from the retina and striate cortex, and intermediate and deep layers that receive inputs from extrastriate, parietal and frontal cortex, and the basal ganglia. Neurons in the superficial layers have responses to visual stimuli that are enhanced when the stimulus is the target of a saccade. Neurons in the intermediate and deep layers may also respond to visual stimuli, and also to auditory stimuli from corresponding locations, so that the visual and auditory maps of the animal's surroundings are in register. It is these intermediate and deep layers that also possess the movement-related activity that plays

BOX 32.2

EYE MOVEMENT DISORDERS

Disorders of eye movements arise from damage to widespread regions of the brain, and the symptoms reflect the contributions of the affected structures to eye motor control.

One common disorder of eye motor control is *strabismus*, which occurs when the movements of the two eyes are not properly yoked. This disorder typically occurs in children or infants (about 2% of children are affected), but can also affect adults. Strabismus is caused by deficits in any of the several factors that are necessary to properly align the eyes, including problems with eye muscles or oculomotor nerves, the presence of a large refractive error in one or both eyes, and lesions in central structures. If strabismus occurs during infancy, while connections in the visual system are developing, it can result in *amblyopia*, in which case visual inputs from the affected eye are mostly ignored.

Many eye movement disorders are caused by tumors, stroke, or other damage to the pathways responsible for forming the eye motor commands. Patients with damage to structures in the brainstem or cerebellum often have difficulty maintaining fixation, and exhibit *nystagmus*—a pattern of involuntary drifts in eye position, interspersed with centering saccades. Nystagmus can be caused by peripheral damage to the vestibular labyrinth or the eighth cranial nerve that transmits signals about head motion to the central nervous system. The imbalance in signals received from the left and right vestibular organs causes a slow drift of the eyes, away from the damaged side. If the nystagmus is caused by peripheral damage, then voluntary saccades and smooth pursuit are mostly unaffected. However, nystagmus can also be caused by central damage to brainstem or cerebellar structures. In these cases, voluntary saccades and smooth pursuit are also impaired, because the functions of the neural integrator and final motor pathways are affected.

Because gaze movements are controlled by descending signals from the cerebral cortex and basal ganglia, eye movement disorders also occur in diseases that affect higher-order brain functions, including schizophrenia, Huntington's disease, and Parkinson's disease. These disorders are characterized by problems with voluntary eye movements, most notably a difficulty in suppressing unwanted saccades during fixation or smooth pursuit eye movements. In Huntington's disease, in addition to difficulty with suppressing unwanted saccades during fixation, patients also have difficulty initiating voluntary saccades. This problem can be partly overcome by blinking at the same time, because the circuits for controlling eye blinks and saccades overlap. In schizophrenia, smooth pursuit eye movements do not correctly match the speed of the moving target, and are interrupted by back-and-forth saccades that repeatedly take the eyes off the target. For reasons that are not yet understood, these deficits in smooth pursuit ability also occur in relatives of schizophrenics who are themselves asymptomatic for the disease.

a crucial role in the control of gaze movements (Wurtz & Goldberg, 1972). These movement-related layers provide a major descending projection to the oculomotor nuclei in the brainstem reticular formation, and also a major ascending projection to the frontal cortex through the medial dorsal nucleus of the thalamus.

The most distinctive feature of the superior colliculus is that it contains a retinotopic map of contralateral visual space and this topography applies to the movement-related activity in the deeper layers as well as to the visual activity in the superficial layers (Fig. 32.10). The significance of this retinotopic map becomes evident when neural activity is artificially manipulated by passing small amounts of electrical current through the tip of an electrode placed in the intermediate layers (Robinson, 1972). Such stimulation elicits a saccadic eye movement with an amplitude and direction that depends on the placement of the electrode in the map. If the stimulation is applied as a long train, rather than as a brief pulse, a series of saccades is elicited, forming a "staircase" of eye movements, each with the same characteristic amplitude and direction. Thus, activity in the superior colliculus map is related to changes in eye position and not to absolute eye position, consistent with the organization of the superior colliculus in retinotopic (or oculocentric) coordinates.

During naturally occurring eye movements, activity in the superior colliculus is not restricted to just the few neurons that best match the endpoint of the saccade, but is distributed across a large number of neurons across the retinotopic map. This point was elegantly demonstrated by testing what happens to saccades when the injection of chemical agents into the superior colliculus temporarily inactivates neurons at a particular location in the map (Lee, Rohrer, & Sparks, 1988). Remarkably, it is still possible to make saccades directed to the center of the inactivated region, because the broad distribution of activity across the map leaves a halo of neurons that can still accurately guide the saccade. Moreover, saccades directed to locations slightly beyond the inactivated site become hypermetric (bigger than normal), because the inactivation removes the contribution from neurons that code for more proximal locations and smaller saccades. These

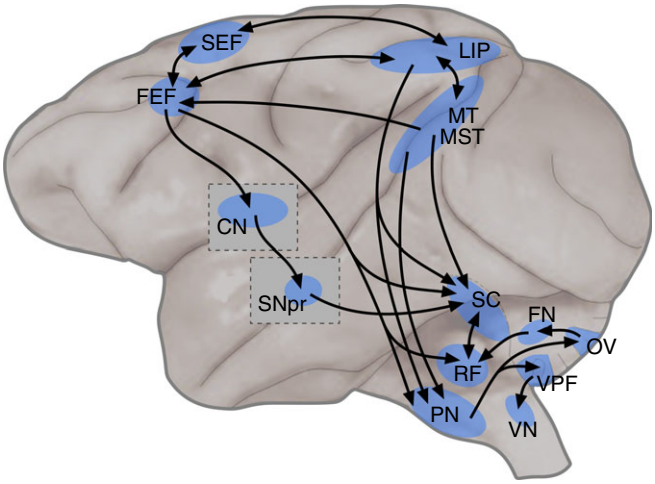


FIGURE 32.9 The descending pathways for controlling gaze movements, depicted schematically on a lateral view of the monkey brain. The superior colliculus (SC), located on the roof of the midbrain, is a major source of descending control signals. In the cerebral cortex, the major areas involved are the frontal eye fields (FEF), supplementary eye fields (SEF), lateral intraparietal area (LIP), middle temporal area (MT), and medial superior temporal area (MST). In the basal ganglia, a cascade through the caudate nucleus (CN) and substantia nigra pars reticulata (SNpr) provides inhibitory control over activity in the superior colliculus. These descending control signals are then converted into motor commands by circuits involving regions such as the reticular formation (RF), pontine nuclei (PN), vestibular nuclei (VN), and parts of the cerebellum such as the ventral paraflocculus (VPF), oculomotor vermis (OV), and fastigial nucleus (FN).

results show that saccades are guided by the vector average of population activity in the superior colliculus, and were the first demonstration of vector averaging in a mammalian motor system.

The intermediate layers of the superior colliculus contain two main classes of neurons related to the control of gaze (Munoz & Wurtz, 1995). Burst neurons in the superior colliculus fire a volley of action potentials that begins ~20 ms before the onset of saccades, somewhat earlier than the burst neurons located downstream in the reticular formation, and often also have a visual response to the onset of visual targets. Unlike the burst neurons in the reticular formation, those in the superior colliculus fire for saccades directed to a portion of the visual field referred to as the neuron's movement field, and the size of the burst depends on where the target falls with respect to the neuron's movement field. Buildup neurons also have visual responses and saccade-related activity, but they are distinguished by the presence of sustained activity when the subject is required to delay their eye movement response after the presentation of a visual stimulus. This sustained activity typically increases over time as the time of the saccade draws near, and is believed to be important for the preparation of saccades, perhaps by facilitating the occurrence of the saccade-related burst. Recent results show that the activity of neurons in these layers of the superior colliculus is not only involved in motor preparation, but is also involved in the antecedent

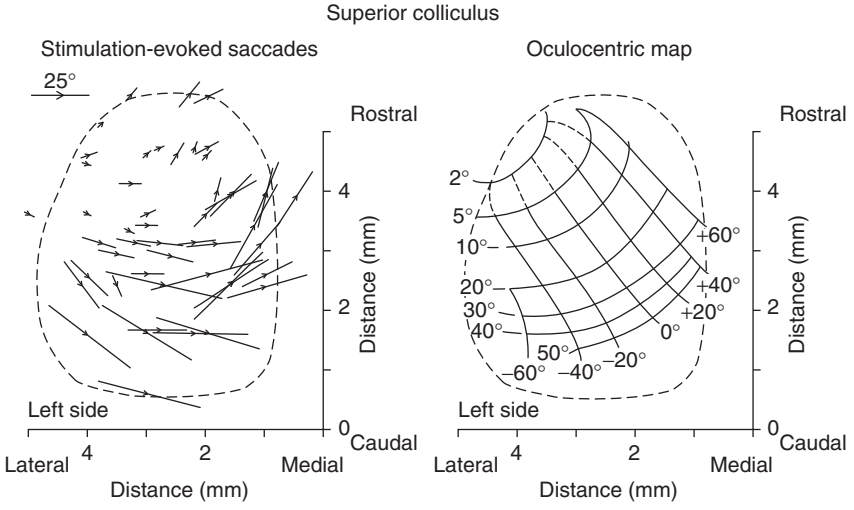


FIGURE 32.10 The superior colliculus and its retinotopic (or oculocentric) map of saccades and target positions. The left plot shows that electrical stimulation in the superior colliculus causes eye movements whose direction and amplitude depend on the location. Each arrow corresponds to a stimulation site in the left superior colliculus, and the length and direction of the arrow indicate the metrics of the evoked saccade. The evoked saccades are all directed rightward, but rostral sites are associated with smaller movements than caudal sites, and medial sites are associated with upward movements whereas lateral sites are associated with downward movements. The right plot summarizes these results in the form of a continuous map of saccades and target positions. As in the visual cortex, the representation of the center of the visual field is amplified. *Adapted from Robinson (1972).*

step of selecting the visual target, and is part of the neural circuits that control visual attention.

The retinotopic map in the superior colliculus includes the central visual field. Like the rest of the map, this part also contains neurons that fire during saccades, although the preferred saccades are smaller, as would be expected from the topography, and even include microsaccades (Hafed, Goffart, & Krauzlis, 2009). However, these same neurons can also show tonic activity during fixation and changes in firing rate for pursuit eye movements. These neurons therefore appear to represent the foveal region of a single "priority" map extending across the superior colliculus that supports multiple eye motor outputs.

The Frontal Eye Field Is the Primary Cortical Area for Controlling Gaze Movements

The frontal eye field (FEF) exerts its control on gaze movements through several descending pathways, including a direct projection to the intermediate layers of the superior colliculus, an indirect projection to the superior colliculus via the basal ganglia, and also projections to the cerebellum and the brainstem reticular formation. The FEF is best known for its role in the generation of saccades, but it also contains zones that are involved in the control of pursuit and vergence eye movements. Like the superior colliculus, electrical stimulation of the FEF causes saccades to a particular location in the contrateral visual hemifield, although the topography in the FEF is less orderly than in the superior colliculus. Electrical stimulation of the pursuit zone of the FEF causes smooth eye movements towards the ipsilateral side, and is the only cortical area where stimulation can directly evoke smooth eye movements. Damage to the FEF produces deficits in saccades, and the impairments are most noticeable when the target stimulus is accompanied by irrelevant distracter stimuli. For pursuit, disruption of the FEF reduces pursuit eye velocity and, in particular, affects the ability of pursuit to follow targets using prediction. The impairments after FEF damage are usually temporary, but combined damage to the FEF and SC results in a permanent deficit in the ability to generate voluntary gaze movements (Schiller, True, & Conway, 1980).

The FEF contains neurons with different combinations of motor and visual activity. Movement-related neurons have movement fields like those found in the superior colliculus; they are active for saccades within a particular range of directions and amplitudes. Movement-related neurons project to the superior colliculus and show changes in activity that are related with when and if voluntary gaze movements will be initiated. In particular, the variability in the timing of saccade initiation is closely correlated with when the firing rates of these neurons reach a fairly constant "threshold" value (Hanes & Schall, 1996). In the pursuit zone of the FEF, neurons exhibit directionally selective responses to moving visual stimuli that are appropriate for determining the velocity command for pursuit eye movements.

Most neurons in the FEF also have visual responses, including a large class of visuo-movement neurons that show activity for both visual stimuli and the eye movements that follow. These visually responsive neurons often show activity related to the selection of the target. Prior to the gaze movement, the activity evoked by the nonselected stimulus is rapidly suppressed, whereas the activity for the selected stimulus remains conspicuous. Some FEF neurons discriminate visual targets even in the complete absence of an eye movement response, indicating that their activity is related to the allocation of visual attention rather than to the preparation of eye movements.

Another region of frontal cortex, the supplementary eye field (SEF), plays a less direct role in the control of saccades but is important for movements that are guided by cognitive factors. Neurons in the SEF are preferentially modulated in tasks that involve more abstract instructions, such as when saccades are directed to a particular side of an object rather than to a spatial location, or when saccades are directed to the location opposite the visual stimulus (the "anti-saccade" task). During pursuit eye movements, SEF neurons are especially active when the target moves along a predictable trajectory.

The Parietal Cortex Contributes to Eye Movements by Controlling Visual Attention

The lateral intraparietal area (LIP) plays a major role in the process of visual selection that precedes the generation of gaze movements. Reversible inactivation of neurons in area LIP does not cause large deficits in the timing or accuracy of gaze movements, but it dramatically reduces the frequency of movements toward the affected portions of the visual field, especially when competing stimuli are present elsewhere (Wardak, Olivier, & Duhamel, 2002). These deficits suggest a visual neglect of stimuli in the affected region, rather than a motor deficit in the generation of eye movements. Damage to the parietal cortex in humans, especially to the right hemisphere, also produces a neglect syndrome with similar deficits in attention.

Neurons in LIP show activity related to visual attention and to the intention to redirect gaze to particular locations in visual space. When an attention-grabbing, but irrelevant, distracter stimulus appears while the animal is preparing a saccade directed elsewhere, the activity of LIP neurons closely tracks the short-lived shift of attention to the distracter and its subsequent return to the intended location of the saccade. Activity in LIP is therefore often described as a salience map that helps to determine the location of spatial attention, as well as the endpoint of gaze movements (Bisley & Goldberg, 2010).

Motion-Processing Areas of Cortex Provide Crucial Signals for Pursuit Eye Movements

The middle temporal area (MT) and the medial superior temporal area (MST) are the primary sources of visual motion information that are used to construct the motor commands for pursuit. This motion information is also used to adjust the size of saccades made to moving visual targets. These cortical areas provide these inputs through several pathways, including projections to other cortical areas involved in controlling gaze movements (in particular, the FEF), descending projections to the vestibulocerebellum via the basal pontine nuclei, and projections to the pretectum, which contains neurons with activity related to the OKR and pursuit.

Neurons in MT and MST have activity related to the velocity of visual stimuli. In MT, neurons have smaller receptive fields and activity that depends on the presence of a visual stimulus. Lesions in MT produce blind spots for seeing motion that affect the motor commands for pursuit and saccades, and the perception of motion as well. In MST, neurons have larger receptive fields and prefer more complex patterns of visual motion. Lesions in MST also produce blind spots for motion, but in addition, cause a directional deficit in pursuit for movements directed to the ipsilateral side (Dürsteler & Wurtz, 1988). Stimulation of MT and MST can affect pursuit eye movements, but only if pursuit is already ongoing, unlike the effects of stimulation in the FEF. Thus, MT and MST are crucial sources of the motion signals needed to drive pursuit eye movements, but other areas appear necessary to determine when and if the movement should be started.

The Basal Ganglia Regulate Saccades by Inhibiting the Superior Colliculus

The caudate nucleus (CN) and the substantia nigra pars reticulata (SNpr) are parts of the basal ganglia that form an important circuit for controlling when gaze movements are initiated. The SNpr applies tonic inhibition to the superior colliculus using GABA-ergic synapses, suppressing the response to the many excitatory inputs to the superior colliculus that might otherwise evoke gaze movements. When SNpr neurons reduce or pause their activity, there is a disinhibition of activity in the superior colliculus, which then permits gaze movements to be initiated. The SNpr itself is under inhibitory control from the CN, which contains neurons that increase their activity with gaze movements. Together, this series of inhibitory connections through the basal ganglia acts to gate the generation of gaze movements by the superior colliculus.

The basal ganglia appear to be especially important for gaze movements involving reward and motivation (Hikosaka, Nakamura, & Nakahara, 2006). For example, the visual responses of CN neurons are larger for stimuli that are associated with larger rewards. This modulation in excitability may be caused by dopaminergic inputs from neurons in the substantia nigra pars compacta (SNpc), the same neurons whose degeneration is implicated in Parkinson's disease.

The Thalamus Provides Feedback about the Control of Gaze Movements

In addition to these descending pathways, there are also ascending pathways to eye movement-related areas of cortex from nuclei in the central thalamus. These nuclei receive inputs from a wide range of oculomotor structures, including the superior colliculus, substantia nigra pars reticulata, cerebellum, the nucleus prepositus hypoglossi, and the brainstem reticular formation. Feedback through the thalamus appears to convey *corollary discharge* signals about the gaze movements that are executed by downstream motor structures, and is used to update the control signals formed in the cortex. Thus, when the central thalamus is inactivated by chemical agents, single saccades can still be made accurately, even in the dark, but subsequent saccades are spatially inaccurate, as though the intervening movement had not been taken into account (Sommer & Wurtz, 2002).

THE CONTROL OF GAZE MOVEMENTS INVOLVES HIGHER-ORDER PROCESSES

Saccades and Pursuit Are Coordinated during Redirecting of Gaze

The functions of saccades and pursuit are distinct, and yet during normal gaze movements the two types of movements are tightly coordinated and almost always select the same visual target. For example, when subjects are presented with two moving stimuli simultaneously, they almost always begin smooth pursuit in a direction that is a weighted average of the two motions. After a brief period of this "vector average" pursuit, subjects make a saccade toward one stimulus or the other and then pursue that one selectively. However, pursuit can selectively follow one stimulus in the presence of distracters without a targeting saccade, showing that saccades are not strictly necessary for pursuit target selection. Pursuit and saccades also exhibit nearly the same tradeoff between speed and accuracy, with saccades tending to be delayed and somewhat more accurate. This suggests that target selection for the two movements is guided by the same decision signals, but saccades use a somewhat more stringent decision criterion (Krauzlis, 2004).

Voluntary Gaze Movements Are Tightly Linked to Shifts of Visual Attention

A crucial aspect of voluntary gaze movements is that they are selectively guided by objects of interest to the

observer, despite the fact that our surroundings usually contain many other distracting signals. This selectivity is achieved by a tight functional link between the control of gaze movements and the mechanisms of visual attention. When gaze is redirected, the movement of the eyes is always preceded by a shift of visual attention to the new location. For example, when subjects are asked to detect visual targets that appear just before a saccade is made, performance is much better when the target and saccade locations coincide than when they do not (Deubel & Schneider, 1996). Similarly during pursuit, subjects make more accurate perceptual judgments about the stimulus they are smoothly tracking with their eyes than about other, nontracked stimuli in the visual display. However, judgments about nontracked stimuli are still better than chance, showing that the selection of targets for gaze movements can occur in parallel with the perceptual processing of other visual stimuli. Also, although redirecting gaze typically requires visual attention, the converse is not true: visual attention can shift to new locations in the absence of gaze movements, and these are referred to as *covert* shifts of attention.

The tight linkage between visual attention and the planning of eye movements was a major rationale for the premotor theory of attention, which posited that the mechanisms for spatial attention and eye motor programming are essentially the same. Although it is generally recognized that the neural mechanisms for attention and eye movements are not identical, there is overlap in their control. Specifically, stimulation within the FEF with currents too weak to directly cause eye movements can nonetheless enhance the visual responses of neurons located in other visual areas of cortex, and can improve the behavioral performance on visual discrimination tasks (Moore & Fallah, 2004). Also, reversible inactivation of the SC causes profound deficits in the performance of spatial attention tasks, even in the absence of targeting eye movements (Lovejoy & Krauzlis, 2010).

Visual Perception and Cognition Contribute to the Control of Gaze Movements

Much of what we know about the neural control of gaze movements was learned using very simple visual, often small bright spots in an otherwise dark room. These studies made it possible to test and identify the basic feedback mechanisms and motor circuits that control gaze. However, under more natural conditions eye movements show properties that are more complex and flexible. One striking set of examples is provided by the work of Yarbus (1967), who showed that the pattern of saccades made while looking at visual scenes depends heavily on the instructions given to the observer (Fig. 32.11). These results illustrate that gaze movements are normally not just reactions to retinal events, but instead are guided by longer-term goals and serve to collect relevant visual information from the environment.

Gaze movements also reveal properties of the visual processing steps required to segment and interpret the visual scene. In one demonstration, a few lights are attached to the rim of a wagon wheel, which is then rolled along in an otherwise dark display. Although the retinal stimulus consists of several spots each undergoing cycloidal motion, subjects perceive the horizontal motion of the wheel and are able to easily track this motion with their eyes (Steinbach, 1976). More recent studies have examined how these processes evolve over time. For many moving objects, the true trajectory of the object cannot be calculated

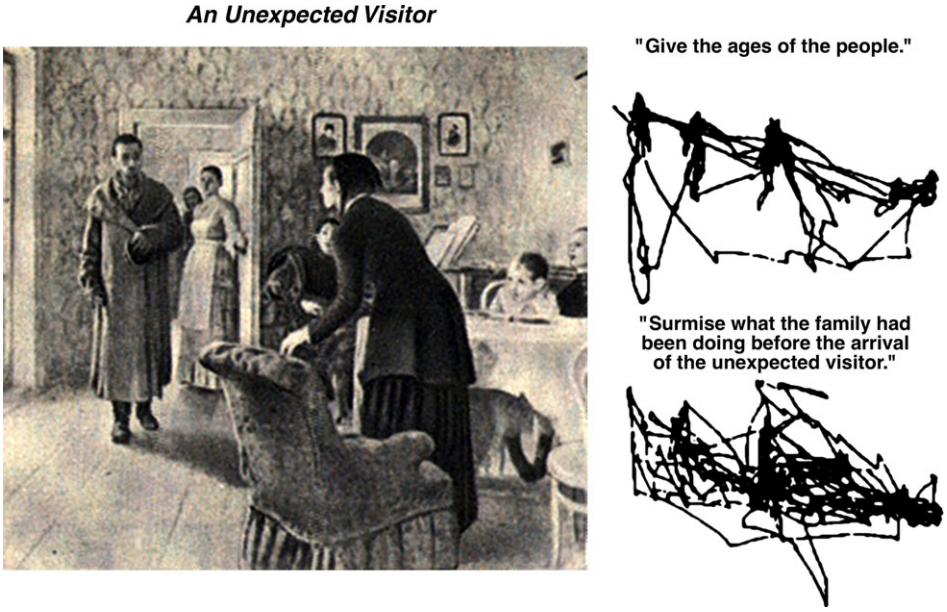

An Unexpected Visitor
"Give the ages of the people."
"Surmise what the family had been doing before the arrival of the unexpected visitor."

FIGURE 32.11 The scan path depends on the goals of the observer. Among other experiments, Yarbus (1967) monitored the eye movements of subjects as they viewed Repin's painting *The Unexpected Visitor* and found that the pattern of eye movements depended on the task. When subjects were instructed to give the ages of the people, the scan path mostly settled on the faces, whereas when subjects were instructed to surmise what the family had been doing, the scan path included many more objects in the room.

locally in the retinal image, but requires integrating information over regions of visual space. This integration evidently takes some time (for example, ~100 ms), and the transition from local-motion signals to object-motion signals can be found in both the eye motor output and perceptual judgments.

Visual perception is also changed during eye movements. In particular, the world is not perceived as moving during saccadic eye movements, despite the rapid motion of the retinal image. You can demonstrate this to yourself by watching yourself make saccades in a mirror: you can see that your eyes change position, but you cannot see them actually move. This perception of a stable world is achieved through a combination of mechanisms, including the use of a *corollary discharge* of the outgoing eye motor command, *saccadic suppression* of visual signals during the eye movement, and *visual masking* by the advent of new visual signals on the retina (Wurtz, 2008). Together, these mechanisms help the visual system reduce its sensitivity to the unreliable visual signals that occur during gaze movements.

CONCLUSIONS

Eye movements serve two different functions. First, eye movements stabilize gaze when animals move about in their environment. Unpredictable high-speed rotations of the head, which are produced whenever an animal moves, change the line of gaze and, by smearing the optical image, reduce the resolution of the visual system. These high-velocity movements are compensated for by counter-rotations of the eyes generated by the VOR. Working in tandem with the VOR is the optokinetic system. The optokinetic system compensates for the slower movements of the head, which happen so gradually that the vestibular system cannot accurately detect them. Second, in animals with retinas that have a small region of high resolution, eye movements redirect gaze to aim the line of sight at objects or features in the visual scene that are of particular interest. Saccades are fast, discrete movements that quickly move the image of a visual target to the center of the retina where acuity is best. Pursuit is a slow, continuous movement that compensates for the motion of the visual target, minimizing the blurring that would otherwise limit acuity. Although the properties of saccades and pursuit are distinct, the two types of movements are closely coordinated during visual search of the environment and both are tightly linked to visual attention. Together, the different types of eye movements make it possible to efficiently gather the visual information that is important for achieving other, longer-term behavioral goals.

Eye movements are controlled by circuits that span the brain. The eyes themselves are a relatively simple motor apparatus, and they are controlled by a final motor pathway in the brainstem that is shared by all classes of eye movements. The construction of the motor commands involves generating an eye velocity command to move the eyes, and an integrated version of this command to hold the eyes at their new position. These motor commands are constructed by circuits in the brainstem reticular formation, and fine-tuned by additional circuits involving several parts of the cerebellum. The descending control signals that drive these motor commands are provided by multiple brain regions, but the superior colliculus in the midbrain and the frontal eye fields in the cortex are especially important. These brain regions help integrate sensory inputs with a variety of higher-order influences so that eye movements are directed to the appropriate visual stimuli. Overall, eye movements provide a window into the brain mechanisms that accomplish motor control, and they also involve sensory-motor interactions and higher-order processes like attention and perception.

References

Bisley, J. W., & Goldberg, M. E. (2010). Attention, intention, and priority in the parietal lobe. *Annual Review of Neuroscience, 33,* 1–21.

Buttner-Ennever, J. A. (2007). Anatomy of the oculomotor system. *Developments in Ophthalmology, 40,* 1–14.

Cannon, S. C., & Robinson, D. A. (1987). Loss of the neural integrator of the oculomotor system from brain stem lesions in monkey. *Journal of Neurophysiology, 57,* 1383–1409.

Demer, J. L. (2006). Current concepts of mechanical and neural factors in ocular motility. *Current Opinion in Neurology, 19,* 4–13.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: evidence for a common attentional mechanism. *Vision Research, 36,* 1827–1837.

Dürsteler, M. R., & Wurtz, R. H. (1988). Pursuit and optokinetic deficits following chemical lesions of cortical areas MT and MST. *Journal of Neurophysiology, 60,* 940–965.

Fuchs, A. F., & Luschei, E. S. (1970). Firing patterns of abducens neurons of alert monkeys in relationship to horizontal eye movement. *Journal of Neurophysiology, 33,* 382–392.

Hafed, Z. M., Goffart, L., & Krauzlis, R. J. (2009). A neural mechanism for microsaccade generation in the primate superior colliculus. *Science, 323,* 940–943.

Hanes, D. P., & Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science, 274,* 427–430.

Hikosaka, O., Nakamura, K., & Nakahara, H. (2006). Basal ganglia orient eyes to reward. *Journal of Neurophysiology, 95,* 567–584.

Keller, E. L. (1974). Participation of medial pontine reticular formation in eye movement generation in monkey. *Journal of Neurophysiology, 37,* 316–332.

Klier, E. M., Meng, H., & Angelaki, D. E. (2011). Revealing the kinematics of the oculomotor plant with tertiary eye positions and ocular counterroll. *Journal of Neurophysiology, 105,* 640–649.

Krauzlis, R. J. (2004). Recasting the smooth pursuit eye movement system. *Journal of Neurophysiology, 91,* 591–603.

Lee, C., Rohrer, W. H., & Sparks, D. L. (1988). Population coding of saccadic eye movements by neurons in the superior colliculus. *Nature, 332,* 357–360.

Lovejoy, L. P., & Krauzlis, R. J. (2010). Inactivation of primate superior colliculus impairs covert selection of signals for perceptual judgments. *Nature Neuroscience, 13,* 261–266.

Miles, F. A. (1997). Visual stabilization of the eyes in primates. *Current Opinion in Neurology, 7,* 867–871.
Missal, M., & Keller, E. L. (2002). Common inhibitory mechanism for saccades and smooth-pursuit eye movements. *Journal of Neurophysiology, 88,* 1880–1892.
Moore, T., & Fallah, M. (2004). Microstimulation of the frontal eye field and its effects on covert spatial attention. *Journal of Neurophysiology, 91,* 152–162.
Munoz, D. P., & Wurtz, R. H. (1995). Saccade-related activity in monkey superior colliculus. I. Characteristics of burst and buildup cells. *Journal of Neurophysiology, 73,* 2313–2333.
Robinson, D. A. (1972). Eye movements evoked by collicular stimulation in the alert monkey. *Vision Research, 12,* 1795–1808.
Schiller, P. H., True, S. D., & Conway, J. L. (1980). Deficits in eye movements following frontal eye field and superior colliculus ablations. *Journal of Neurophysiology, 44,* 1175–1189.
Sommer, M. A., & Wurtz, R. H. (2002). A pathway in primate brain for internal monitoring of movements. *Science, 296,* 1480–1482.
Steinbach, M. (1976). Pursuing the perceptual rather than the retinal stimulus. *Vision Research, 16,* 1371–1376.
Wardak, C., Olivier, E., & Duhamel, J. R. (2002). Saccadic target selection deficits after lateral intraparietal area inactivation in monkeys. *Journal of Neuroscience, 22,* 9877–9884.
Wurtz, R. H., & Goldberg, M. E. (1972). Activity of superior colliculus in behaving monkey: III. Cells discharging before eye movements. *Journal of Neurophysiology, 35,* 575–586.
Wurtz, R. H. (2008). Neuronal mechanisms of visual stability. *Vision Research, 48,* 2070–2089.
Yarbus, A. L. (1967). *Eye movements and vision.* New York: Plenum.

Suggested Readings

Angelaki, D. E., & Hess, B. J. M. (2005). Self-motion-induced eye movements: effects on visual acuity and navigation. *Nature Reviews Neuroscience, 6,* 966–976.
Fuchs, A. F., Kaneko, C. R. S., & Scudder, C. A. (1985). Brainstem control of saccadic eye movements. *Annual Review of Neuroscience, 8,* 307–337.
Gandhi, N., & Katnani, H. A. (2011). Motor functions of the superior colliculus. *Annual Review of Neuroscience, 34,* 205–231.
Kowler, E. (2011). Eye movements: the past 25 years. *Vision Research, 51,* 1457–1483.
Krauzlis, R. J. (2005). The control of voluntary eye movements: new perspectives. *Neuroscientist, 11,* 124–137.
Leigh, R. J., & Zee, D. S. (2006). *The neurology of eye movements* (4th ed.). Philadelphia, PA: F.A. Davis.
Lisberger, S. G. (2010). Visual guidance of smooth-pursuit eye movements: sensation, action, and what happens in between. *Neuron, 66,* 477–491.
Purcell, B. A., Heitz, R. P., Cohen, J. Y., Schall, J. D., Logan, G. D., & Palmeri, T. J. (2010). Neurally constrained modeling of perceptual decision making. *Psychological Review, 117,* 1113–1143.
Schutz, A. C., Braun, D. I., & Gegenfurtner, K. R. (2011). Eye movements and perception: a selective review. *Journal of Vision, 11,* 1–30.
Sommer, M. A., & Wurtz, R. H. (2008). Brain circuits for the internal monitoring of movements. *Annual Review of Neuroscience, 31,* 317–338.
Sparks, D. L. (2002). The brainstem control of saccadic eye movements. *Nature Reviews Neuroscience, 3,* 952–964.

Richard J. Krauzlis

SECTION VI

REGULATORY SYSTEMS

This page intentionally left blank

CHAPTER

33

The Hypothalamus: An Overview of Regulatory Systems

The hypothalamus is an integrative center essential for survival of an organism and reproduction of its species. Regulatory systems emerged as each organism adapted to its environment and have evolved to control the complex interactions of physiology and behavior.

The diversity in regulatory function of the hypothalamus is reflected in its structure, neurochemistry, and connections. Almost every major subdivision of the central nervous system (CNS) communicates with the hypothalamus and is subject to its influence. In addition, the hypothalamus communicates with peripheral organ systems by converting synaptic information to blood-borne hormonal signals. In turn, the hypothalamus responds to feedback from the peripheral systems that it regulates. The chapters in this section discuss the processes by which the hypothalamus influences physiology and behavior, and the purpose of this chapter is to provide an introductory context for those presentations. We begin by providing a historical perspective on research that has defined the hypothalamus as an important regulatory center, continue with an overview of its structural and functional organization, and conclude with a discussion of how this region of the diencephalon functions within brain circuitry devoted to the control of motivated behaviors.

HISTORICAL PERSPECTIVE

Research on the hypothalamus has a rich and storied history. Although depicted in early brain atlases, the hypothalamus was not recognized as a distinct division of the diencephalon until the work of Wilhelm His (1893). Our understanding of the structure and function of the hypothalamus developed as basic and clinical neurosciences were integrated. Harvey Cushing, a neurosurgeon who studied endocrine disorders from a neuroscience perspective, provided important and influential early insights into hypothalamic function. He described diabetes insipidus, a neuroendocrine dysfunction of uncontrolled urinary water excretion and thirst, and identified the importance of injury to the pituitary stalk in the condition. Cushing also described the syndrome that bears his name, Cushing's syndrome, a disorder characterized by excessive secretion of adrenal corticoids, and he was among the first to recognize that the hypothalamus and pituitary form a functional unit. He was a pioneer in the development of pituitary surgery and is viewed as a founder of the fields of neurosurgery and endocrinology. In 1929, summing up his work in this area shortly before his retirement, Cushing wrote of the hypothalamus, "*Here in this well-concealed spot, almost to be covered with a thumb nail, lies the very mainspring of primitive existenceve getative, emotional, reproductiveon which, with more or less success, man has come to superimpose a cortex of inhibitions.*"

Gurdijian (1927), Kreig (1932), and Le Gros Clark, Beattie, Riddoch, and Dott (1938) build upon the clinical observations of Cushing and others, using histological procedures to describe the cytoarchitecture of hypothalamic nuclei. Their observations proved to be remarkably predictive of functional specializations revealed by the work of Ranson, Hess, and others during the same period. These investigators demonstrated that lesions and localized stimulation of the hypothalamus could selectively affect appetite, weight control, water balance, autonomic control, reproductive function, and emotional behavior. This work progressed so well that in 1940 it could be summarized only in a large volume of the annual series of the "Association for Research in Nervous and Mental Disease." In that volume, Scharrer and Scharrer (1940) presented evidence for "neurosecretory neurons," providing

© 2013 Elsevier Inc. All rights reserved.

the foundation for a fundamental tenet of hypothalamic function. That is, there are neurons—in the brain—that secrete hormones directly into the bloodstream. In 1949, Wolfgang Bargmann demonstrated neurosecretory neurons with cell bodies in the hypothalamus and axons projecting to the posterior lobe of the pituitary gland. Bargmann's work provided a framework for our current understanding of neuroendocrine regulation, that is, release of peptides into the peripheral circulation allows the hypothalamus to exert a profound influence on the physiology of peripheral organs.

In the 1950s, Harris and Green defined the means through which the hypothalamus regulates the anterior lobe of the pituitary. Their identification of a neurovascular link, the portal plexus, through which the hypothalamus regulates the anterior pituitary, fundamentally altered concepts of hypothalamic function and neuroendocrine regulation. The neurochemical basis of anterior pituitary regulation was revealed in the 1970s when Guillemin and Schally independently isolated and characterized peptide hormones that acted on the anterior pituitary. In subsequent years the field of neuroendocrinology built enormously on these observations, and the resulting literature documents a diverse family of hypothalamic peptides that are transported through the portal plexus to stimulate or inhibit the release of other hormones from the anterior lobe of the pituitary gland.

Contemporary research has exploited a revolution in technology development to bring further transparency to the organization and function of the hypothalamus. Data derived from a variety of new technical approaches has dramatically improved our understanding of the hypothalamic "connectome," including the phenotype of constituent neurons, and the underlying mechanisms that allow integrated, adaptive regulation of physiology and behavior. Similarly, advances in molecular technology have allowed targeted functional dissection of the diverse regulatory systems resident in the hypothalamus. While much remains to be learned, our knowledge of hypothalamic regulatory systems has advanced considerably in a relatively short period of time.

We have made great progress in improving understanding of Cushing's "mainspring of primitive existence." Fundamental insights into CNS function have emerged from the analysis of hypothalamic structure and function, underscoring the importance of this small, complex subdivision of diencephalon in the regulation of homeostasis and behavior essential to survival.

HYPOTHALAMIC CYTOARCHITECTURE

Early studies of hypothalamic organization defined subdivisions that were largely based upon landmarks visible on the ventral surface of the brain. Three major subdivisions were defined based upon their associations with the optic chiasm (chiasmatic), the protuberance containing the pituitary stalk (tuberal), and the mammillary bodies (mammillary) (Fig. 33.1). Contemporary research indicates

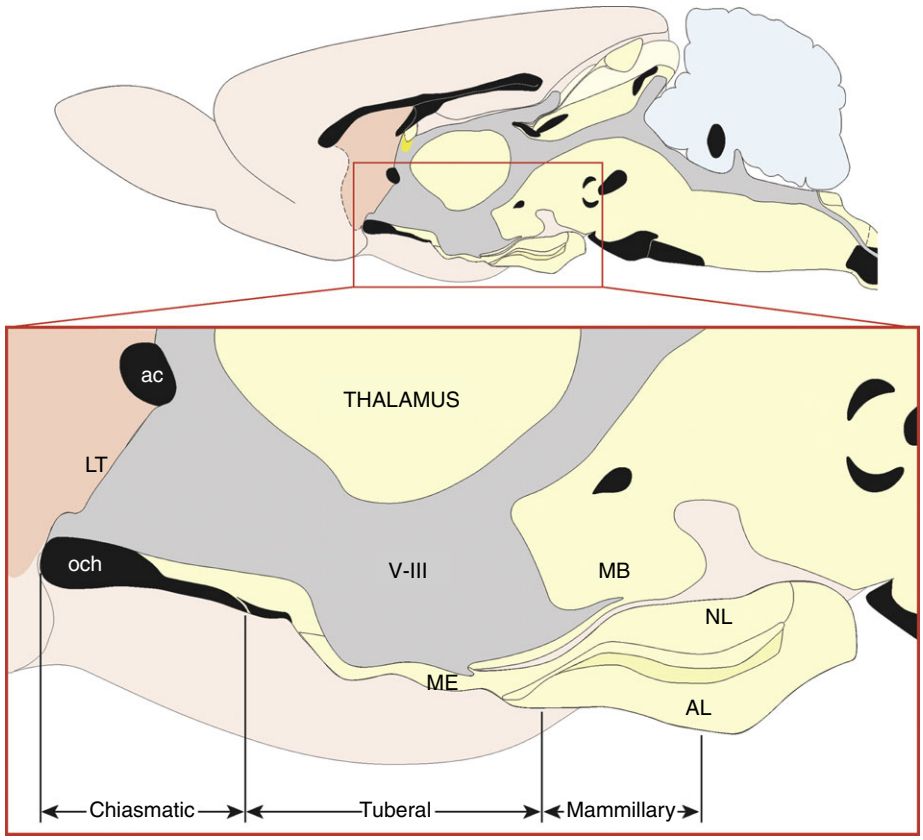


FIGURE 33.1 Boundaries of the rat hypothalamus viewed in midsagittal section. The ventricular system is shaded gray. The boxed area shows the approximate location of the hypothalamus in the ventral quadrant of the diencephalon, below the thalamus. The floor of the third ventricle (V–III), formed by the optic chiasm (och), median eminence (ME), and rostral portion of the mammillary body (MB), defines a large portion of the rostrocaudal extent of the hypothalamus. These landmarks also define the major subdivisions (chiasmatic, tuberal, and mammillary) classically used to subdivide the hypothalamus. The tuberal hypothalamus is distinguished by the infundibular stalk, which connects the ventral diencephalon to the pituitary. The rostral limit of the hypothalamus is the lamina terminalis (LT), a thin strip of tissue extending between the optic chasm and anterior commissure (ac) that also forms the rostral wall of the third ventricle. NL, neural lobe of posterior pituitary; AL, anterior lobe of pituitary.

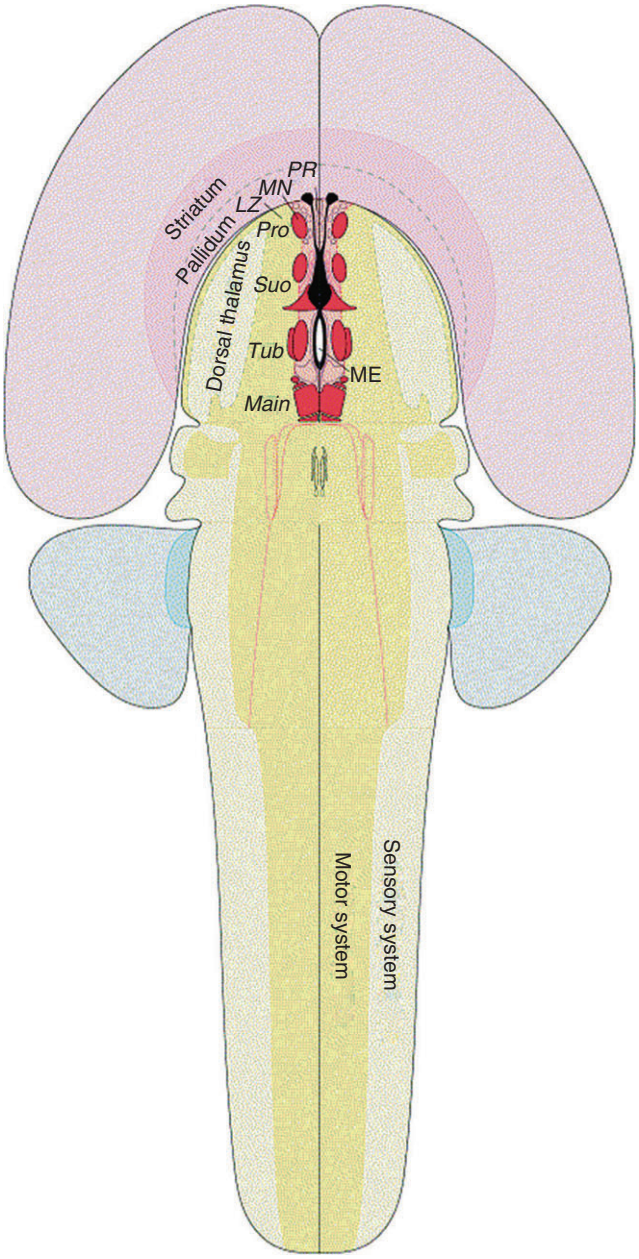


FIGURE 33.2 The columnar organization of hypothalamic cell groups is illustrated on a flat map of the rat central nervous system. Three longitudinal zones form columns that extend through the rostrocaudal extent of the hypothalamus. The periventricular region (PR; light red) lies immediately adjacent to the third ventricle and contains cell groups responsible for neuroendocrine control of the pituitary gland (black). The medial nuclei consist of cytoarchitecturally distinct cell groups (dark red) between the periventricular and lateral zones. The diffusely organized but phenotypically distinct neurons of the lateral zone (LZ) are shown in yellow. Contemporary research also recognizes four transverse subdivisions of hypothalamus (pro, preoptic; supraoptic or anterior, SUO; tuberal, tub; mammillary, mam). *Source: Reprinted with permission from Swanson (2000).*

that the hypothalamus is organized into longitudinal columns on either side of the third ventricle (Fig. 33.2). These columns are largely consistent with the initial descriptions of Gurdjian (1927) and Krieg (1932) and consist of *periventricular*, *medial*, and *lateral* zones that extend through the entire rostrocaudal extent of the hypothalamus. The periventricular zone, as the name implies, lies adjacent to the ependymal lining of the third ventricle. Its constituent neurons are densely packed within a narrow zone adjacent to the ependyma and occasionally expand into cytoarchitecturally distinct nuclear groups such as the suprachiasmatic and arcuate nuclei. The medial zone lies adjacent to the periventricular zone and contains cell groups that vary in size and cell morphology, but are organized into cytoarchitecturally distinct nuclei. This contrasts with the diffusely distributed neurons of the lateral zone, which is the least distinct of the three columns. Neurons in all three regions have been further subdivided on the basis of cellular phenotype, connectivity, and cellular markers of neural activation, allowing investigators to impose functional definitions.

FUNCTIONAL ORGANIZATION OF THE HYPOTHALAMUS

To effectively integrate diverse physiological processes and behaviors, the hypothalamus must communicate with systems in both the brain and periphery. Projections from all major subdivisions of the nervous system terminate within the hypothalamus. In some instances, such as vision, monosynaptic projections bring first-order sensory information from the periphery directly to hypothalamic nuclei. In other cases, such as limbic projections, multisynaptic processed information is relayed to the hypothalamus. More than any other region of the nervous system, the hypothalamus depends on conversion of synaptic information to hormonal signals. In addition, the hypothalamus is subject to feedback regulation by the physiological processes that it controls. These diverse systems are best understood by looking at functions that they modulate.

Olfaction

Multisynaptic pathways allow the olfactory system to influence neurons in many parts of the hypothalamus. Olfactory information passes directly to the olfactory tubercle, piriform cortex, and amygdala, which in turn relay it to the hypothalamus through the medial forebrain bundle (corticohypothalamic projections), stria terminalis, ventral amygdalofugal pathway, and fornix. In many vertebrates, olfaction is integral to behaviors involved in reproduction, defense, and feeding, and this is reflected in the large amount of forebrain dedicated to the processing of olfactory stimuli. Insight into the complexity of these circuits is apparent in recent rodent studies employing transneuronal tracing technology to define polysynaptic olfactory circuits innervating hypothalamic neurons involved in the regulation of reproduction and fertility (Boehm, Zou, & Buck, 2005; Yoon, Enquist, & Dulac, 2005). Thus, olfaction is an important sensory modality integral to hypothalamic functions essential to reproduction and survival.

Visual Projections

A principal function of the hypothalamus is to impose temporal organization on hormone secretion and behavior through the timekeeping properties of the biological clock in the rostral hypothalamus. The neurons that constitute this clock are found within the suprachiasmatic nuclei (SCN). The genetically determined circadian rhythmicity of SCN neurons provides a timekeeping capacity that is essential to successful adaptation to the environment and, in many instances, reproduction of organisms. Importantly, the rhythmic activity of the biological clock is responsive to environmental cues, which is accomplished through central visual projections of the retina. A unique population of ganglion cells that contain light sensitive melanopsin transduces light in the retina. The central projections of these ganglion cells through the retinohypothalamic projection synchronize the circadian activity of SCN neurons with the daily cycle of light and dark. Polysynaptic connections from the clock influence melatonin secretion from the pineal gland as part of a feedback mechanism for measuring day length, an ability important for the timing of reproduction in photoperiodic species. The responsiveness of this regulatory system to variations in photoperiod is also responsible for the "jet lag" that humans suffer when traveling across time zones.

Visceral Sensation

Sensory information reaches the hypothalamus through ascending projections arising from neurons in the nucleus of the solitary tract (NTS) of the caudal brain stem. The NTS is the principal visceral sensory nucleus that receives topographically organized input from the major organ systems of the body by way of cranial nerves X and IX, the vagus and glossopharyngeal nerves. As such, it is the first region in the CNS to process information about visceral, cardiovascular, and respiratory functions, as well as taste. The NTS coordinates reflex modulation of peripheral organ function and sends processed sensory information to forebrain nuclei, where it is integrated to control more complex physiological processes and behaviors.

Neurons in the paraventricular hypothalamic nucleus (PVH) and the lateral hypothalamic area (LHA) are prominent among a subset of hypothalamic nuclei that receive direct projections from the caudal brainstem through the medial forebrain bundle. In addition, the NTS and parabrachial nuclei relay splanchnic visceral and nociceptive sensory information from the spinal cord. This "interoceptive" sensory feedback has an important influence upon autonomic and endocrine outputs of the hypothalamus essential for regulation of homeostasis, and also influences mood through interactions with higher brain regions. Evidence for the latter is particularly prevalent in literature examining sensory integration related to stress induced activation of the endocrine and autonomic axes.

Integration of Physiology and Behavior

In the 1930s, James Papez (1937) postulated that the mammillary nuclei in caudal hypothalamus were components of a multisynaptic pathway, classically known as Papez's circuit, responsible for the expression of emotion. Although this function has not been confirmed, contemporary studies have demonstrated that projections, direct and polysynaptic, to hypothalamus from higher brain centers, play an integral role in coordinating adaptive changes in physiology with affect. Advances in defining the identity and organization of projection systems influential in this regulation have substantially advanced our understanding of how this is achieved. It is now well established that the prefrontal cortex, hippocampus, and amygdala contribute to this regulation through extensive topographically organized projections to the hypothalamus.

Prominent among the observations that have forged this vision are studies demonstrating that the hippocampal formation influences the hypothalamus through both direct and indirect projections through the fornix. The direct path, or postcommissural fornix, is a prominent component of Papez's original circuit in that it is composed of axons that arise from neurons in the hippocampal formation and terminate in the mammillary bodies. The indirect path, or "precommissural," fornix, influences the hypothalamus through a synaptic relay in the septum (Risold & Swanson, 1996). The topography of this projection is apparent at its origin in the hippocampus and is preserved in the lateral septal nuclei and in the hypothalamus, where second order septal projections innervate three longitudinally organized hypothalamic columns, although not equally (Fig. 33.3). Thus, this projection system is in a position to regulate neuroendocrine and autonomic function and ingestive behavior (periventricular zone), modulate motivated reproductive and agonistic behaviors (rostral medial zone), and also provide polysynaptic feedback projections to the hippocampus via the mammillary complex. The demonstration of "place" and "head-direction" cells in the hippocampal formation suggests that this projection is also the source of adaptive multimodal sensory information relevant to a variety of hypothalamic functions (e.g., foraging).

Projections from the amygdala to the hypothalamus and cortex coordinate endocrine and autonomic responses with the behavioral expression of fear. Projections from the amygdala reach the hypothalamus through two prominent projection pathways. The stria terminalis arises from neurons in the amygdala and follows a looping trajectory similar to the course of the fornix. Axons coursing through this pathway ramify within the bed nuclei of stria terminalis (BST) and continue into the hypothalamus. Amygdala neurons also innervate the hypothalamus via the ventral amygdalofugal bundle, which passes directly over the optic tract into the hypothalamus.

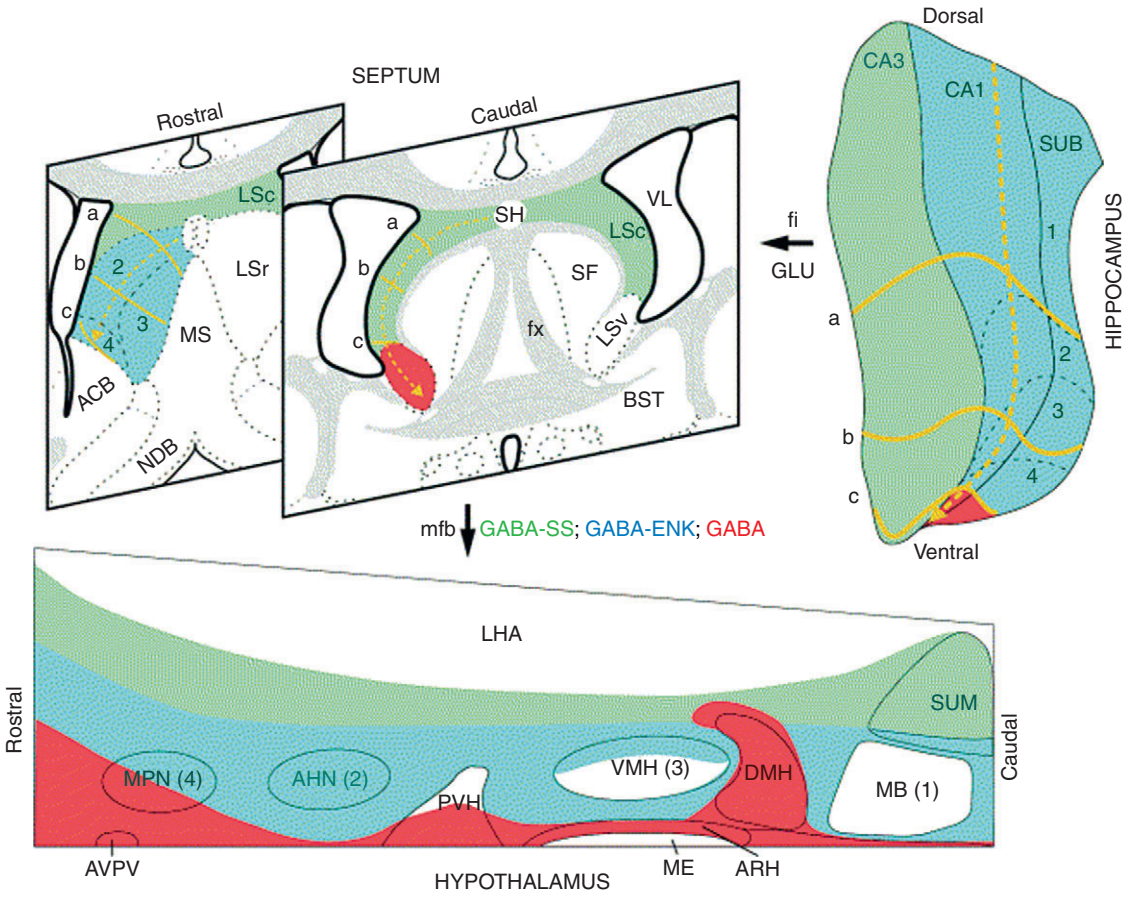


FIGURE 33.3 The organization of pathways that bring hippocampal input to hypothalamic columns is illustrated. The hippocampal formation projects topographically upon the lateral subdivision of the septum which, in turn, projects topographically upon the periventricular (red), medial (blue), and lateral (green) zones of the hypothalamus. Subfields of the hippocampal formation (CA1, CA3, subiculum) are represented on the flat map of the unfolded hippocampus in the upper right. The topography of the projection systems is coded in the colors, numbers (1, 2, 3, 4), and small letters (a, b, c) included in the maps of hippocampus, septum, and hypothalamus. ACB, nucleus accumbens; AHN, anterior hypothalamic nucleus; ARH, arcuate nucleus; AVPV, anteroventral periventricular nucleus; BST, bed nucleus of stria terminalis; DMH, dorsomedial nucleus; ENK, enkephalin; fi, fimbria of fornix; fx, fornix; GLU, glutamate; LS, lateral septum (small letters designate rostral, caudal, and ventral subdivisions); MB, mammillary body; ME, median eminence; mfb, medial forebrain bundle; MPN, medial preoptic nucleus; MS, medial septum; NDB, nucleus of the diagonal band; PVH, paraventricular nucleus; SH, septohippocampal nucleus; SS, somatostatin; SUM, supramammillary nucleus; VL, lateral ventricle; VMH, ventromedial nucleus. *Source: Reprinted with permission from Risold and Swanson (1996).*

A large literature has established the BST, hippocampal formation, and medial prefrontal cortex, and paraventricular nucleus of thalamus (PVT) as essential components of the neural circuitry through which the brain responds to stressful stimuli (Hsu & Price, 2009; Ulrich-Lai & Herman, 2009). The BST and PVT are integral to this regulation in that it integrates information from a variety of systems in response to psychogenic and physical stressors and exerts control over the endocrine (hypothalamic-pituitary-adrenal axis) and autonomic regulatory centers in hypothalamus. The paraventricular nucleus of hypothalamus (PVH) is a particularly important nodal point in this control in that it houses neurons at the apex of the HPA axis and also influences autonomic function through descending projections.

Circumventricular Organs

Not all sensory information reaching the hypothalamus is of synaptic origin. Some regions of the CNS are chemosensitive to blood-borne molecules. Such information about body fluid composition has important implications for the maintenance of homeostasis. Circumventricular organs (CVOs) are integral to this regulation. Seven CVOs surround the ventricular system in the diencephalon, midbrain, and hindbrain (Fig. 33.4). These include the subfornical organ (SFO), the vascular organ of the lamina terminalis (OVLT), the median eminence (ME), the pituitary gland (PG), the pineal gland (P), the subcommissural organ (SCO), and the area postrema (AP) (Fig. 33.3). Three of these regions, the OVLT, ME, and PG, are components

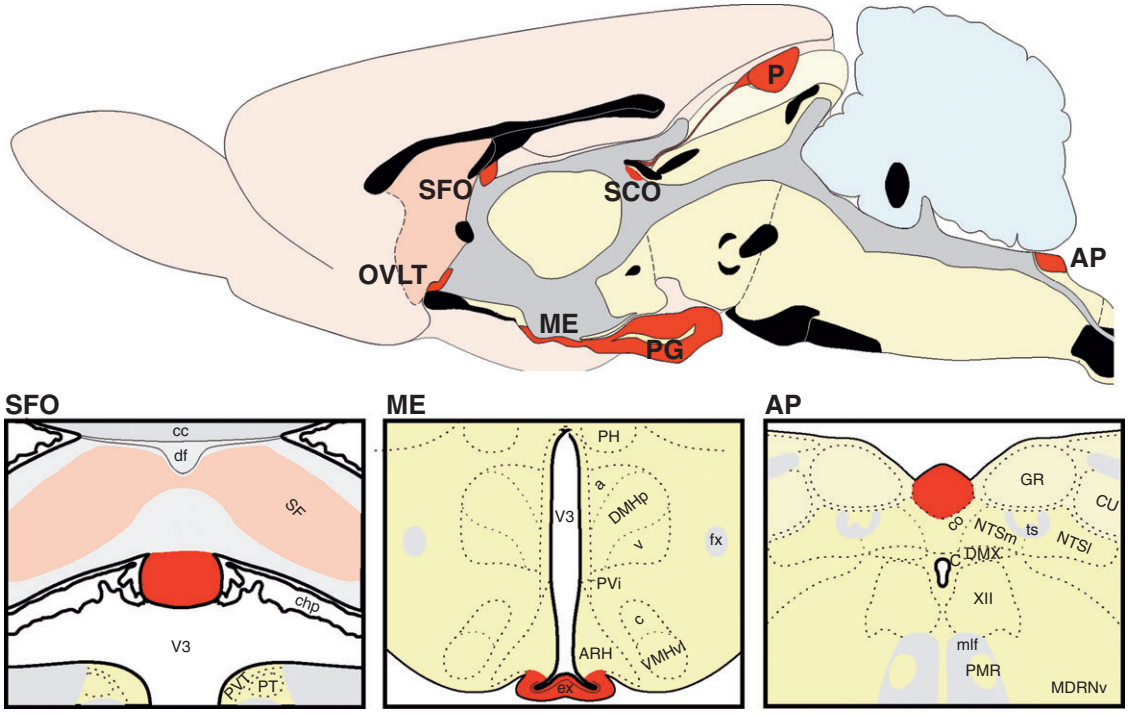


FIGURE 33.4 The location of the seven circumventricular organs (shown in red) in the rat brain (midsaggital section) are illustrated. Three CVOs that have an intimate functional association with the hypothalamus are illustrated in transverse sections in the lower figures. AP, area postrema; ARH, arcuate nucleus; cc, corpus callosum; CU, cuneate nucleus; df, dorsal fornix; DMHa, anterior portion of the dorsomedial nucleus; DMHp, posterior portion of dorsomedial nucleus; DMHv, ventral portion of dorsomedial nucleus; DMX, dorsal motor vagal nucleus; GR, gracile nucleus; ME, median eminence; mlf, medial longitudinal fasiculus; co, commissural portion of the nucleus of the solitary tract; NTSl, lateral portion of the nucleus of the solitary tract; NTSm, medial portion of the nucleus of the solitary tract; OVLT, vascular organ of the lamina terminalis; PH, posterior hypothalamus; P, pineal gland; PMR, paramedian reticular nucleus; PVi, intermediate part of periventricular nucleus; SCO, subcommissural organ; SF, septofimbrial nucleus; SFO, subfornical organ; ts, tractus solitarius; V–III, third ventricle; VMHc, central part of ventromedial nucleus; VMHvl, ventrolateral part of ventromedial nucleus; XII, hypoglossal nucleus.

of the hypothalamus. Two others, the SFO and AP, have extensive connections with hypothalamic nuclei involved in neuroendocrine and homeostatic function. Paul Ehrlich identified these regions in the late 1800s when he noted that they selectively accumulated vital dyes injected into the peripheral vasculature. Later studies demonstrated that the dye accumulated due to fenestrations in the capillaries in these regions that permitted relatively large molecules to leave the vascular lumen and enter the extracellular milieu. As a result, CVO neurons are accessible to blood-borne molecules that cannot enter other regions of the brain. The absence of a blood–brain barrier also allows CVOs to influence peripheral function through the release of bioactive molecules. Thus, CVOs represent important "windows" through which the brain controls peripheral function and, importantly, that render CNS neurons responsive to feedback regulation by the secretory products of the target organs that they modulate. This type of architecture is fundamental to the way in which the hypothalamus achieves tightly regulated control over the pituitary and, by extension, its peripheral targets.

EFFECTOR SYSTEMS OF THE HYPOTHALAMUS ARE HORMONAL AND SYNAPTIC

The extensive influence of the hypothalamus on behavior and physiology is reflected in its effector systems. The hypothalamus has conventional synaptic connections with essentially every major subdivision of the CNS and plays a major role in the control of behavioral state. As noted earlier, it also has regulatory control over the pituitary gland and thereby exerts a profound influence over peripheral physiology. Thus, the output of hypothalamic systems is integral to the integrated expression of physiology and behavior.

Behavioral State Control

Regulation of the sleep–wake cycle is a distributed process that involves the integrated action of the reticular activating system and the diencephalon. Early studies employing lesions and knife cuts conducted by Moruzzi,

Magoun, and Bremer demonstrated the importance of the brain stem reticular activating system in the generation of consciousness. Subsequently, Nauta (1946) postulated the existence of sleep centers in the hypothalamus based upon the consequences of lesions to the rostral or caudal hypothalamus. In those rodent studies, he demonstrated that lesions of the preoptic region produced insomnia and that somnolence resulted from lesions of the caudal hypothalamus.

Until recently, little was known regarding the specific hypothalamic cell groups involved in sleep regulation. However, identification of the circadian pacemaker in the early 1970s by Moore and colleagues marked the beginning of an era in which the hypothalamic contributions to behavioral state regulation have become increasingly clear. This literature has confirmed the early hypotheses of Nauta and identified cell groups in the rostral and caudal hypothalamus whose activity correlates with sleep or arousal.

The rostral hypothalamus is resident to two cell groups that are influential in the regulation of the sleep–wake cycle. Of these, the suprachiasmatic nuclei, or "Mind's Clock," are well known for their genetically determined oscillatory activity. A large literature has established that SCN activity opposes the homeostatic drive for sleep and is essential for the temporal organization of rest and activity (Chapter 40). In contrast, Saper, Scammell, and Lu (2005) identified hypothalamic neurons that become active at the onset of sleep in rats. These cells, located in the ventrolateral preoptic region, contain the inhibitory neurotransmitter GABA and project to the caudal hypothalamic cell groups that produce arousal when activated. These observations reveal a circuitry whose function coincides with the opposing effects of the rostral and caudal hypothalamus in sleep regulation documented by Nauta.

Two distinct populations in caudal hypothalamus generate arousal when activated. These neurons are similar in that they project diffusely to the cortical mantle but differ in phenotype. One group uses histamine as a neurotransmitter and is resident in the tuberomammillary nuclei. Another group is present in the posterior and lateral hypothalamus and contains the neuropeptide known as hypocretin (or orexin). Studies of the causal mechanisms underlying the sleep disorder known as narcolepsy implicate this population of neurons in the regulation of behavioral state. Narcolepsy is a disorder characterized by excessive daytime sleepiness and sudden inappropriate intrusion of sleep during waking. A defect in one of the hypocretin receptors has been demonstrated in a canine model widely employed to study the disease, and transgenic mice lacking the hypocretin gene display sleep–wake disturbances similar to narcolepsy. Additionally, cerebrospinal fluid (CSF) levels of hypocretin are reduced in humans suffering from this disorder, and postmortem analysis of the hypothalamus of narcoleptics has demonstrated substantial reductions in the number of hypocretin neurons.

Collectively, these findings are consistent with an important role for the hypothalamus in the regulation of behavioral state. Nevertheless, it is important to emphasize that the hypothalamus functions within a larger ensemble of neurons devoted to behavioral state regulation. Particularly important in this regard are monoaminergic and cholinergic brain stem neurons, as well as thalamic and basal forebrain cell groups (Chapter 40). Understanding how this distributed system functions in an integrated fashion to control behavioral state remains an important goal.

Neuroendocrine Regulation through the Pituitary Gland

The median eminence is the gateway through which the hypothalamus exerts regulatory control over peripheral organ systems. This CVO is continuous with the pituitary gland through a thin tissue bridge known as the pituitary stalk. Parvicellular hypothalamic neurons exercise this regulatory control over the anterior lobe of the pituitary via the release of peptides and small molecule neurotransmitters into fenestrated portal vessels that connect the median eminence with another fenestrated plexus in the anterior pituitary. These molecules act upon cells in the anterior pituitary to control release of hormones into the peripheral circulation. The posterior lobe of the pituitary gland also contains a fenestrated vascular plexus that is contacted by axons of magnocellular neurons that project through the median eminence and pituitary stalk to release their peptides—oxytocin or vasopressin—into the vasculature. These peptides are well known for their influences on fluid homeostasis (vasopressin) and milk letdown in lactating females (oxytocin).

The parvicellular hypothalamic neurons acting in the anterior lobe of the pituitary exert a profound and diverse influence over peripheral physiology and metabolism. A variety of parvicellular neuroendocrine neurons distinguished by peptide phenotype contribute to this regulation either by stimulating or by inhibiting the release of hormones from the anterior lobe of the pituitary. The PVH and arcuate nuclei contain a large proportion of these neurons. As noted previously, the PVH is the apex of the HPA axis, the neuroendocrine arm through which the brain responds to stressful stimuli. Parvicellular neurons that secrete corticotropin-releasing hormone provide the essential neuroendocrine output of the hypothalamus that activates this axis. PVH parvicellular neurons that synthesize and secrete thyrotropin-releasing hormone are similarly influential in the regulation of peripheral metabolism. Considered collectively, the parvicellular neurons in the PVH, arcuate nucleus, and other regions of the

hypothalamus provide the means through which the brain regulates peripheral systems essential to the life of the organism and the survival of species. The integrative capacity of the hypothalamus therefore is essential to the regulation of the many systems that operate through this neuroendocrine "final common pathway."

Reproductive Function

The ability of mammals to reproduce depends on the function of the hypothalamus and its ability to communicate with the pituitary gland. This is especially true for the production of gametes, whose maturation and release in both males and females require a functionally intact hypothalamic–pituitary axis. The hypothalamus also plays an essential role in the organization and expression of the complex behaviors necessary for copulation (see Chapter 38).

Experimental studies in rodents have demonstrated that the rostral medial zone of the hypothalamus plays an important role in orchestrating reproductive behaviors. The medial preoptic region participates in the control of masculine sexual behavior, such as erection, mounting, and ejaculation, whereas ventral regions of the hypothalamus play a major role in the control of feminine sexual behaviors, such as lordosis.

A number of hypothalamic cell groups implicated in reproductive function are sexually dimorphic (Simerly, 2002). For example, ultrastructural studies of a preoptic region in male and female rats have demonstrated differences in synaptology that develop in response to perinatal exposure to androgens. Also, differences in the size and serotonergic innervation of the medial preoptic nuclei of male and female rats have been demonstrated. The possibility that similar dimorphisms are present in the human hypothalamus is supported by the identification of a region of the human preoptic region that is approximately twice as large in males as in females. Although the functional significance of sexual dimorphism in the structure and connectivity of the hypothalamus remains to be established, the gender of the brain is clearly determined by the hormonal milieu present during a critical period of development.

Much is known about the hypothalamic cell groups that influence the production of gametes. The production of ova and sperm is controlled by the release of gonadotropins from the anterior pituitary gland, which in turn is regulated by the hypothalamus in a gender-specific manner. For example, spermatogenesis occurs in response to constant release of gonadotropins in males, whereas ovulation is a cyclic event initiated by a rhythmic surge in gonadotropin release in females. In both cases, hypothalamic neurons produce gonadotropin-releasing hormone (GnRH) to stimulate the release of gonadotropin from the anterior lobe of the pituitary. The gender-specific patterns of gonadotropin release imply that there are sexual differences in the functional organization of circuitry that controls the activity of GnRH neurons. One postulated sexually dimorphic difference is a central pattern generator, thought to be responsible for the cyclic release of GnRH in females. The identity, organization, and regulation of the cell groups that contribute to this pattern generator remain an active and important focus of investigation.

Central Integration of Autonomic Function

The hypothalamus is a prominent component of the central network of neurons that regulates autonomic function. Descending projections from the hypothalamus to autonomic cell groups in the brain stem and spinal cord exert a strong influence upon the sympathetic (thoracic) and parasympathetic (cranial-lumbosacral) divisions of the autonomic nervous system (ANS). Several hypothalamic cell groups also indirectly influence the ANS through relays that integrate processed sensory information and project to the preganglionic neurons of the spinal cord that are the final common pathway for autonomic outflow. Brain stem neurons involved in cardiovascular regulation are among the best-characterized examples of such relays. For example, brain stem neurons in the rostroventrolateral medulla (RVLM) exert control over mean arterial pressure through direct projections to sympathetic preganglionic neurons in the intermediolateral cell column of thoracic spinal cord (Guyenet, 2006). The activity of these neurons is controlled in a reflex manner through a caudal brain stem circuit that integrates baroreceptor sensory input terminating in the NTS (Chapter 34). However, descending projections from the hypothalamus (e.g., paraventricular nucleus) and other CNS cell groups important for coordinating autonomic function and behavioral responses to environmental challenges also influence the activity of RVLM neurons. This is one of many relays through which the central network devoted to autonomic control provides a dynamic regulatory capacity ideally suited for maintaining homeostasis and ensuring adaptive changes in physiology to environmental challenges.

Immune Function and Thermoregulation

A major function of the hypothalamus is to integrate sensory information and elicit adaptive changes in physiology and behavior to achieve homeostasis. As noted above, this is accomplished through both synaptic interactions with other regions of the neuraxis and neuroendocrine regulation of peripheral systems. Immune activation is a major challenge to homeostasis so it should not be surprising that there is communication between the immune system and the hypothalamus. Clear documentation that such interactions

exist has largely emerged from contemporary literature that has benefited from the identification and characterization of pro- and anti-inflammatory cytokines and analyzed their effect upon neural systems. The seminal contributions of Hugo Besedovksy have been particularly influential in establishing that immune activation alters the turnover of classical neurotransmitters (e.g., noradrenaline) and activates the HPA axis. Numerous investigators have built upon these observations and a large literature has established that interactions between the immune and nervous systems are bidirectional; for example, immune activation can alter neuronal function and the brain also has the capacity to elicit changes in the immune response.

Fever is one consequence of an activated immune system. Thus, the neural circuits that control temperature regulation provide important insights into neural-immune interactions. In this regard, it is well known that the hypothalamus contains populations of neurons that either stimulate or inhibit heat production. These include heat sensitive neurons in the medial preoptic area that act to reduce core body temperature by inhibiting hypothalamic and brainstem cell groups that generate heat production. Nakamura and Morrison (2008) recently brought further insight into the way in which this circuitry functions to defend body temperature. These investigators demonstrated that sensory information activated by skin cooling is relayed to a "thermoregulatory command center" in the preoptic hypothalamus through the parabrachial nucleus to induce compensatory thermogenic, metabolic, and cardiovascular responses. This sensory transduction occurs in parallel with somatosensory pathways that mediate temperature perception and discrimination through thalamocortical projections. In rodents and other small mammals, nonshivering thermogenesis is achieved through the activation of sympathetic outflow to brown adipose tissue (BAT). The mitochondria of BAT adipocytes contain an uncoupling protein that converts energy from fatty acid metabolism to heat, a process that is increased by sympathetic activity. Interestingly, there is considerable overlap of cell groups that regulate energy balance, stress, and fever. Although there is much to be learned regarding the neural systems involved in temperature regulation, it is clear that the hypothalamus is at the center of the sensory integration necessary to respond to interoceptive challenges (e.g., inflammatory responses to infection) as well as those encountered in the environment.

Fluid Homeostasis and Thirst

Insight into the regions of the brain involved in fluid homeostasis has come from observing the response of neurons to dehydration and other experimental changes in the internal milieu. The hypothalamus is a prominent component of a set of CNS nuclei in which neurons respond to dehydration (an increase in fluid osmolarity) and changes in blood volume by increasing their metabolic activity. Activated neurons have been demonstrated in the subfornical organ, the paraventricular and supraoptic nuclei, and an area of the rostral hypothalamus known as the AV3V region, which includes the median preoptic nucleus and the OVLT. These regions interact with neurons in the area postrema, the nucleus of the solitary tract, the noradrenergic cell groups of the brain stem, and the parabrachial nucleus to integrate peripheral stimuli and elicit adaptive changes in the fluid environment of the organism. CVOs subserve prominent regulatory functions within this circuitry (Fig. 33.5).

The SFO has dense projections to the supraoptic and paraventricular nuclei, to the region adjacent to the OVLT, and to the perifornical region of the lateral hypothalamic area. Such projections are consistent with postulated roles of the SFO in autonomic and homeostatic functions. The peptide hormone angiotensin II (AII) is integral to the way in which the SFO transduces peripheral signals and then produces compensatory neural responses. When blood pressure falls, the kidneys release renin into the bloodstream. Renin triggers a biochemical cascade that produces AII. Because of the absence of a blood–brain barrier in the SFO, circulating AII can enter the SFO and bind AII receptors on SFO neurons. Interestingly, SFO neurons appear to use AII as a neurotransmitter in projections to hypothalamic cells that are involved in the control of fluid dynamics. In this example, activation of hypothalamic

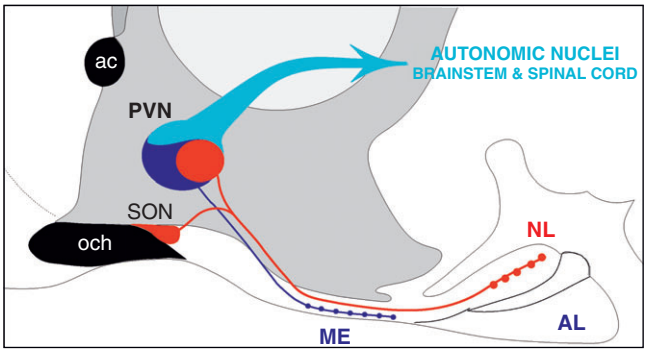


FIGURE 33.5 Functional associations of the paraventricular nucleus (PVH) with endocrine and autonomic systems are illustrated. Axons of magnocellular neurons in the PVH and supraoptic nuclei (SON) (shown in red) traverse the internal layer of the median eminence (ME) and the infundibular stalk to terminate on a vascular plexus in the neural (posterior) lobe of the pituitary gland (NL). Parvicellular PVH neurons influence the functional activity of the anterior lobe of the pituitary gland (AL) through projections (shown in blue) that terminate on a fenestrated capillary plexus, the portal plexus, in the external zone of the median eminence. The terminals of these axons release peptides and neurotransmitters into the portal vasculature, which carries them to the anterior lobe of the pituitary. Parvicellular PVH neurons also give rise to descending projections (shown in turquoise) to autonomic nuclei in the brain stem and spinal cord. ac, anterior commissure; och, optic chiasm.

systems that elicit drinking and reduce fluid excretion at the kidneys would produce the adaptive compensatory homeostatic increase in blood pressure. Further details on the components of the regulatory system are provided in Chapter 37.

Different populations of hypothalamic neurons elicit adaptive responses to changes in fluid osmolarity. Osmosensitive neurons are located primarily in the AV3V region, which includes the OVLT. The response of these neurons to alterations in fluid osmolarity elicits an adaptive compensatory thirst. The neural circuitry that modulates this response is not fully understood but is known to involve input from the OVLT, SFO, and NTS. Further, it takes advantage of the same output regulatory pathway, hypothalamic magnocellular neurons, utilized by the SFO. Chapter 37 considers the hypothalamic role in fluid homeostasis in greater detail.

Food Intake

The consumption of food is another complex behavior that involves several regions of the brain. The hypothalamus plays an essential role in this circuitry as an integrative center for a variety of sensory stimuli, both synaptic and hormonal. It also serves to modulate autonomic outflow to the viscera. This modern view of neural regulation of food intake is a substantial modification of the simple concept of opposing centers for hunger and satiety that was popular in the older literature. Chapter 36 contains a detailed discussion of literature that has contributed to the modern view of the neural control of food intake. The following description summarizes the functional organization of hypothalamic nuclei that regulate food intake.

When food is consumed, hormones are released from the viscera, and sensory pathways innervating the alimentary tract are stimulated. These sensory pathways regulate neurons in the area postrema and NTS that are connected bidirectionally with hypothalamic nuclei. Hypothalamic nuclei are also subject to feedback regulation from hormonal signals arising in the gut. At least four hypothalamic nuclei are involved in the control of feeding, and a prominent part of their influence is exerted via descending projections to autonomic nuclei in the brain stem and spinal cord. Paraventricular nuclei provide a prominent descending hypothalamic influence. In particular, parvicellular oxytocinergic neurons in the PVH provide an important descending projection to the dorsal vagal complex (including the NTS) and the sympathetic preganglionic neurons of the thoracic spinal cord. At least some of this descending influence appears to be under the control of neuropeptide Y (NPY)-containing neurons that project from the arcuate nucleus to the PVH. In experiments, infusion of NPY into the PVH stimulates food intake and NPY-containing neurons in the arcuate nucleus are inhibited by insulin released into the peripheral circulation after a meal and by leptin, a peptide released from adipose tissue. Thus, hypothalamic nuclei regulate food intake through the autonomic nervous system, and a subset of these neurons is influenced by peripheral signals from the periphery (see Chapter 36).

Summary

It is clear that the hypothalamus plays a pivotal regulatory role central to the survival and propagation of vertebrate species. This regulation is achieved by output pathways that utilize both synaptic and hormonal mechanisms and is subject to feedback regulation by sensory information derived from the systems that they regulate. The integrative capacities of the hypothalamus are fundamental to its ability to orchestrate the complex and diverse adaptive behaviors that underlie homeostasis, reproduction of the species, and motivated behaviors. Within this context it is important to view the hypothalamus as an integral component of ensembles of neurons devoted to specific functions rather than an individual unit. Swanson has advanced a conceptual framework consistent with this philosophy. In this formulation, the hypothalamus is part of a "behavioral control column" that functions as a topographically organized integrative center subject to the control of cerebral hemispheres. The functional organization of this column is consistent with the distributed localization of functions classically associated with the hypothalamus. For example, the rostral portion of the column is devoted to the control of ingestive (eating and drinking) and social (defensive and reproductive) behaviors, whereas the caudal portion, which is continuous with the substantia nigra, functions in support of exploration and foraging. Modulation of these functions is under the control of the cerebral hemispheres through three descending projections involving the cortex, striatum, and pallidum. The role of the hypothalamus in imparting temporal organization on the behavioral state and physiology is an important component of this organizational framework. The explosion of organizational and functional data that has accumulated over recent decades has expanded our understanding of hypothalamic function and established this small but complex subdivision of the diencephalon as an important integrative unit for the expression of behavior and maintenance of homeostasis.

References

Boehm, U., Zou, Z., & Buck, L. B. (2005). Feedback loops link odor and pheromone signaling with reproduction. *Cell, 123,* 683–695.

Gurdijian, E. S. (1927). The diencephalon of the albino rat. *Journal of Comparative Neurology, 43,* 1–114.

Guyenet, P. G. (2006). The sympathetic control of blood pressure. *Nature Reviews Neuroscience, 7,* 335–346.

His, W. (1893). Proposals for subdividing the brain (Vorschlage zur Einteilung des Gehirns). *Archiv filr Anatomie und Entwicklungsgeschichte* (Leipzig), *17,* 157–171.

Hsu, D. H., & Price, J. L. (2009). Paraventricular thalamic nucleus: Subcortical connections and innervation by serotonin, orexin, and corticotropin-releasing hormone in macaque monkey. *Journal of Comparative Neurology, 512,* 825–848.

Krieg, W. J. S. (1932). The hypothalamus of the albino rat. *Journal of Comparative Neurology, 55,* 19–89.

Le Gros Clark, W. E., Beattie, W. E., Riddoch, W. E., & Dott, N. M. (1938). *The hypothalamus*. Edinburgh: Oliver & Boyd.

Nakamura, K., & Morrison, S. F. (2008). A thermosensory pathway that controls body temperature. *Nature Neuroscience, 11,* 62–71.

Nauta, W. J. H. (1946). Hypothalamic regulation of sleep in rats: An experimental study. *Journal of Neurophysiology, 9,* 285.

Papez, J. W. (1937). A proposed mechanism of emotion. *Archives of Neurology and Psychiatry, 38,* 725–743.

Risold, P. Y., & Swanson, L. W. (1996). Structural evidence for functional domains in the rat hippocampus. *Science, 272,* 1484–1486.

Saper, C. B., Scammell, T. E., & Lu, J. (2005). Hypothalamic regulation of sleep and circadian rhythms. *Nature, 437,* 1257–1263.

Scharrer, B., & Scharrer, E. (1940). Sensory cells within the hypothalamus. In J. F. Fulton, S. W. Ranson & A. M. Frantz (Eds.), *The hypothalamus* (pp. 170–194). Baltimore: Williams & Wilkins.

Simerly, R. B. (2002). Wired for reproduction: Organization and development of sexually dimorphic circuits in the mammalian forebrain. *Annual Review of Neuroscience, 25,* 507–536.

Swanson, L. W. (2000). Cerebral hemisphere regulation of motivated behavior. *Brain Research, 886,* 113–164.

Ulrich-Lai, Y. M., & Herman, J. P. (2009). Neural regulation of endocrine and autonomic stress responses. *Nature Reviews Neuroscience, 10,* 397–409.

Yoon, H., Enquist, L. W., & Dulac, C. (2005). Olfactory inputs to hypothalamic neurons controlling reproduction and fertility. *Cell, 123,* 669–682.

Suggested Readings

Ader, R. (1996). Historical perspectives on psychoneuroimmunology. In H. Friedman, T. W. Klein & A. L. Friedman (Eds.), *Psychoneuroimmunology, stress, and infection* (pp. 1–24). Boca Raton, FL: CRC Press.

Ehrlich, P. (1956). Uber die methylenblaureaction der lebenden Nervensubstanz. In F. Himmelweit (Ed.), *The collected papers of Paul Ehrlich* (Vol. 1, pp. 500–508). London, England: Pergamon Press.

Fulton, J. F., Ranson, S. W., & Frantz, A. M. (Eds.), (1940). *The hypothalamus*. Baltimore: Williams & Wilkins.

Galoyan, A., & Besedovsky, H. (Eds.), (2008). *Handbook of neurochemistry and molecular neurobiology, neuroimmunology*. (3rd ed.). New York: Springer.

Harris, G. W. (1948). Neural control of the pituitary gland. *Physiological Reviews, 28,* 139–179.

Hess, W. R. (1957). *The functional organization of the diencephalon*. New York, NY: Grune & Stratton.

Johnson, A. K., & Gross, P. M. (1993). Sensory circumventricular organs and brain homeostatic pathways. *FASEB Journal, 7,* 678–686.

Rinaman, L. (2007). Visceral sensory inputs to the endocrine hypothalamus. *Frontiers in Neuroendocrinology, 28,* 50–60.

Saper, C. B. (1990). Hypothalamus. In G. Paxinos (Ed.), *The human nervous system* (pp. 389–414). San Diego: Academic Press.

Swanson, L. W., & Cowan, W. M. (1975). Hippocampo-hypothalamic connections—Origin in subicular cortex, not Ammons Horn. *Science, 189,* 303–304.

J. Patrick Card and Larry W. Swanson

This page intentionally left blank

CHAPTER

34

Central Control of Autonomic Functions: Organization of the Autonomic Nervous System

The autonomic nervous system (ANS) consists of the neural circuitry that controls the body's physiology. This network is a morphologically, embryologically, functionally, and pharmacologically distinct division of the nervous system. Working in concert with neuroendocrine mechanisms, the ANS is responsible for homeostasis, or what Walter B. Cannon (1939) referred to as the "wisdom of the body."

Unlike the skeletal motor system, which innervates striated muscles, the autonomic nervous system innervates the body's smooth muscle organs and tissues. The ANS projects to, and forms plexuses within, the hollow organs, or viscera, such as the heart and lungs in the thorax and the gastrointestinal, genital, and urinary tracts in the abdomen. It also projects to the blood vessels, glands, and other target tissues found within trunk muscles, limb muscles, and skin.

The ANS also coordinates the body's physiology to support behavior. In tandem with endocrine controls, the ANS orchestrates the continuous adjustments in blood chemistry, respiration, circulation, digestion, reproductive status, thermoregulation, and immune responses that protect the integrity of the internal milieu, while enabling the functioning and coordination of skeletal muscles and exteroceptive senses into what we know as behavior. Autonomic adjustments are often phasic responses that occur with latencies typical of neural reflexes; in contrast, endocrine responses usually occur more slowly, taking anywhere from minutes to hours to seasons for full expression.

Autonomic impairments lead to the loss of ability to mobilize physiological responses to counter challenges to homeostasis. For example, Cannon (1939), in his landmark experiments, found that destruction of different autonomic pathways devastated the ability of an animal to regulate body temperature, to respond to perturbations of fluid and electrolyte balance, to control blood sugar homeostasis, or to mobilize in response to threats.

A familiar experience illustrates autonomic operations: occasionally, when you have been lying down resting and then jump up suddenly, you may feel dizzy, your vision may become blurred, and you may momentarily have trouble maintaining your balance—let alone doing something active and coordinated. This phenomenon is postural, or orthostatic, hypotension and is caused by insufficient oxygenated blood reaching your brain under the altered hemodynamic conditions produced by the rapid change in posture. Most people experience postural hypotension only occasionally, though for some it is a chronic medical problem. Indeed, it is a debilitating condition in people with various forms of dysautonomia (Bannister, 1989).

A principle underscored by postural hypotension is that behaviors, indeed even simple movements such as a change in posture, require near-instantaneous redistributions of blood flow throughout the body to protect the privileged supply of blood, and thus oxygen and glucose, to the brain. The demands of postural adjustments require unceasing monitoring and rebalancing of circulatory loads. The fact that these adjustments happen continuously and automatically is testimony to the efficiency of the ANS in producing homeostatic adjustments. This example also illustrates two other characteristics of many autonomic responses: often the responses either are initiated in anticipation of the perturbation or

© 2013 Elsevier Inc. All rights reserved.

implemented so rapidly that the individual does not experience a deficit, and the activity of the ANS is coordinated with activity of the somatic nervous system.

The scope of autonomic functions is hard to overemphasize. Though many early experiments concentrated on autonomic emergency responses, ANS operations are by no means limited to cardiovascular and metabolic adjustments, to "fight-or-flight" situations, or to short-term responses. The ANS has equally essential roles, for example, in all reproductive behavior and physiology, including the choreographies of copulation, gestation, parturition, and lactation. Autonomic reflexes control, on a continuous basis, all bodily interfaces with the environment (e.g., the dermal, respiratory, and digestive epithelia). By its rapid and unceasing adjustments of pupillary diameter and lens accommodation, the ANS facilitates optimal vision. And the ANS plays critical roles in immune system responses; it also coordinates the elimination of metabolic waste products.

The autonomous nature of ANS function is another characteristic of autonomic control. We are rarely aware of ongoing reflex adjustments made to maintain cardiovascular and regulatory dynamics, fluid and electrolyte homeostasis, energy balance, immune system operations, or reproductive functions. Like the batch files of a computer, which execute without normally intruding on the screen or requiring management, autonomic reflex adjustments run unceasingly in the "background." This makes it practical for other sensory events and functional decisions to occupy the "foreground." In one of his delightful and trenchant essays, Lewis Thomas (1974) speculates on how disastrous and overwhelmingly confusing it would be if he were put in charge of his liver—if he had to consider all visceral inputs and make a conscious decision for all adjustments in liver function. The ANS takes care of such hepatic, as well as all other visceral, decisions automatically, allowing the individual to focus on the behavioral and cognitive functions that typically require awareness.

The hub of the autonomic nervous system is the visceral motor outflow, which is divided into *sympathetic* (SNS) and *parasympathetic* (PSNS) divisions. Each division is organized hierarchically into pre- and postganglionic levels (Fig. 34.1). The cell bodies of the preganglionic neurons lie within the central nervous system, specifically in the brainstem and spinal cord. Descending projections from more rostral levels of the neuroaxis converge on these ANS preganglionic motor neurons in a pattern similar to that of the centrifugal projections to somatic motor neurons of the brainstem and spinal cord. However, unlike the outflow pattern of the skeletal motor system, in which neurons of the ventral horn or the cranial nerve nuclei project directly to the effector or muscle target, autonomic preganglionic neurons project to peripheral ganglia located between the last CNS station and the target tissues. These ganglia of the ANS contain

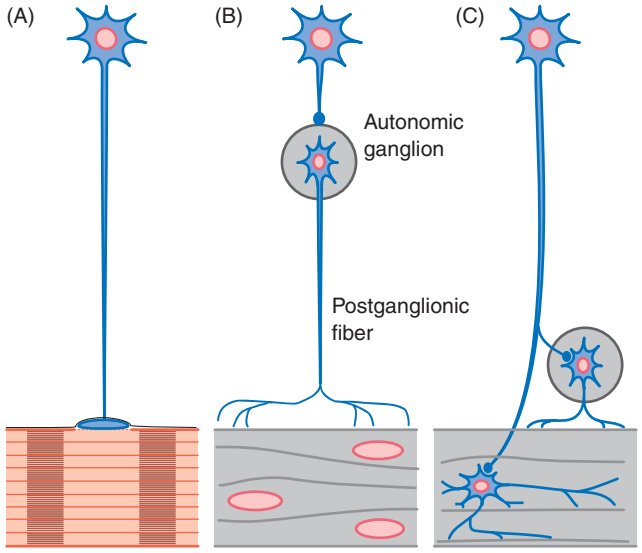


FIGURE 34.1 Somatic and autonomic styles of motor innervation are different. In the somatic motor pattern (A), motor neurons of the spinal cord and cranial nerve nuclei project directly to striated muscles to form neuromuscular junctions. In the autonomic or visceral pattern (B and C), in contrast, motor neurons in the central nervous system project to peripheral postganglionic neurons that in turn innervate smooth muscles. (B) The SNS has preganglionic neurons in the thoracic and lumbar spinal cord (intermediolateral cell column) that project to postganglionic motor neurons in para- and prevertebral autonomic ganglia. (C) The PSNS consists of preganglionic neurons in brainstem cranial nerve nuclei and in the sacral spinal cord (intermediomedial cell column) that project to postganglionic motor neurons in ganglia located near or inside the viscera. *From Nauta and Feirtag (1986).*

the postganglionic motor neurons that project to the effector tissues. The extra synapse interrupting the autonomic outflows at peripheral ganglia allows for more divergence, and it also provides opportunities for more local integrative circuitry to impinge on the outflow. Different locations both of the preganglionic motor neurons in the CNS and of the ganglia containing the somata of the postganglionic motor neurons in the periphery are distinguishing features of the two principal divisions of the autonomic nervous system (Langley, 1921; Table 34.1).

SYMPATHETIC DIVISION: ORGANIZED TO MOBILIZE THE BODY FOR ACTIVITY

Sympathetic Responses Coordinate Energy Expenditure, Catabolic Functions, and Adjustments for Intense Activity

"Fight or flight," Cannon's prototypical example, epitomizes the operation of the sympathetic division of the ANS. Say you are walking alone at night in an unfamiliar part of town. Although the area is somewhat unsafe and it is too dark to see much of your surroundings, the night is quiet, and you manage to relax. After you walk for several

TABLE 34.1 Classical Comparisons of Sympathetic and Parasympathetic Divisions of the ANS[a]

Characteristic	SNS	PNS
Location of Preganglionic Somata	Thoracolumbar Cord	Brainstem; Sacral Cord
Location of Ganglia (and thus Postganglionic Somata)	Distant from Target Tissues	Near or in Target Tissues
Postganglionic Transmitter	Norepinepherine	Acetylcholine
Length of Preganglionic Axon	Relatively Short	Relatively Long
Length of Postganglionic Axon	Relatively Long	Relatively Short
Divergence of Preganglionic Projections	One-to-Many	One-to-Few
Functions	Catabolic	Anabolic
Innervates Trunk and Limbs, in addition to Viscera	Yes	No

[a]*Langley (1921), in splitting the ANS motor outflow into the now-accepted divisions, stressed the contrasting and distinguishing traits of the sympathetic (SNS) and parasympathetic (PSNS) divisions of the ANS. Since his time, most discussions have identified several contrasting features. Most of these distinctions apply generally, but not universally. Although many of these distinctions have blurred (e.g., transmitter) or may be completely unfounded (e.g., divergence), they are often discussed.*

minutes, however, there is a sudden, loud, and unfamiliar noise close by and behind you. In literally the time of a heartbeat or two, your physiology moves into high gear. Your heart races; your blood pressure rises. Blood vessels in your skeletal muscles dilate, increasing the flow of oxygen and energy. At the same time, blood vessels in your gastrointestinal tract and skin constrict, making more blood available to be shunted to skeletal muscle. Digestion is inhibited; release of glucose from the liver is facilitated. Your pupils dilate, improving vision. You begin to sweat, a response serving several functions, including promoting additional dissipation of heat so muscles can work efficiently. Multiple other smooth and cardiac muscle adjustments occur automatically, and almost all of them are affected by the sympathetic division of the ANS.

This fight-or-flight example illustrates important features of the ANS generally and its SNS division more particularly. First, the situation points out the utility of the housekeeping functions of the ANS. If the alarming noise turns out to be a threat and it is necessary to fight or flee, skeletal muscles must be optimally tuned and provisioned. Extra quantities of oxygen and energy are essential. Second, the synergy of adjustments indicates a coordinated and adaptive program of responses, which in this case happen immediately, almost before cognitive evaluation. Third, the short latency of the response is characteristic. Some of the physiological adjustments could be accomplished by hormonal mechanisms, but such adjustments would be too slow to be of much immediate help.

Though the flight-or-fight scenario highlights short-latency responses, not all sympathetic (or parasympathetic) responses happen on such short time scales. ANS operations also involve continuous, ongoing adjustments. Textbook accounts of the emergency functions of the ANS and of single, isolated reflex adjustments sometimes leave the incorrect impression that the ANS is quiescent in the absence of a threat or specific stimulus. Instead, the autonomic nervous system, like the somatic nervous system, always maintains an operating tone. Even during rest, the system maintains appropriate homeostatic balances and autonomic programs.

Preganglionic Neurons of the Sympathetic Nervous System Are Located in the Thoracic and Lumbar Spinal Cord

Sympathetic preganglionic neurons occupy the *intermediolateral nucleus*, a columnar grouping of cells running longitudinally, in the lateral horn of the spinal gray (Strack, Sawyer, Marubio, & Loewy, 1988). In humans, these cells are found between the first thoracic spinal segment (T1) and the third lumbar segment (L3) (Fig. 34.2A). The location of these preganglionic neurons and their axonal outflow gives the SNS another of its names: the *thoracolumbar division*. Cells within the intermediolateral column are aggregated segmentally. There is a rostral-to-caudal viscerotopic organization to the distribution (Fig. 34.2A). The column is distributed bilaterally, and preganglionics on one side send their axons out the ipsilateral ventral root. The axon of a preganglionic neuron typically exits from the segment in which its soma is located. Many of the preganglionic axons are lightly myelinated, and, as they separate from the ventral root to project to the appropriate peripheral ganglion, they form a *white ramus*, a connective that takes its descriptive name from myelinated fibers.

Sympathetic postganglionic neurons are found in two distinct types of peripheral ganglia: paravertebral and prevertebral. *Paravertebral ganglia*, as the name implies, are adjacent to the spinal cord bilaterally, in a position slightly ventral and lateral to the vertebral column (Figs. 34.3 and 34.4). Each ganglion receives a white ramus from the appropriate ventral root. These ganglia also are interconnected longitudinally into a sympathetic paravertebral chain linked by axons from the preganglionic neurons that run rostrally or caudally to neighboring ganglia of the chain. Axons of the postganglionic neurons of the paravertebral ganglia course centrifugally out of the ganglion to join the peripheral nerves that will carry the individual fibers to their targets. Because most postganglionic axons in the sympathetic nervous system are unmyelinated, these connectives joining the ventral roots are called *grey rami*.

Paravertebral ganglia of the sympathetic chain supply the postganglionic innervation of the head, thorax, trunk,

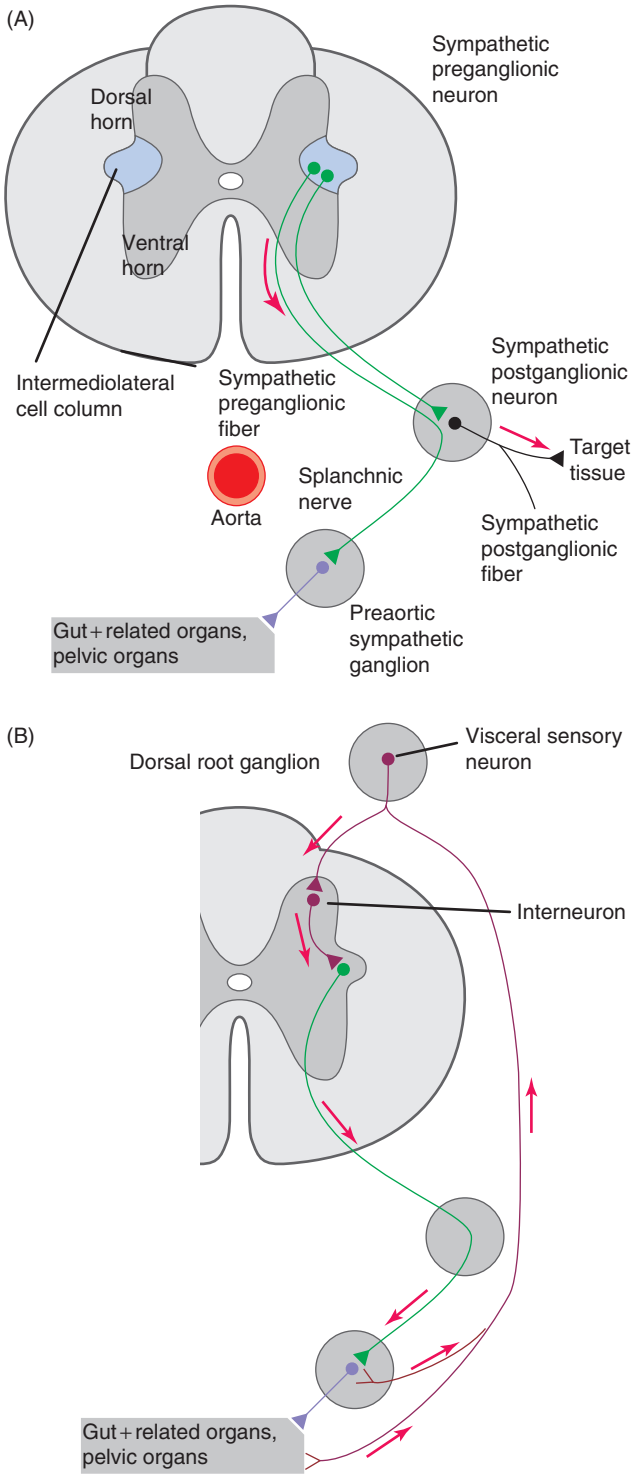


FIGURE 34.2 Details of the organization of the SNS and its sensory inputs. In thoracic and lumbar levels of the spinal cord, preganglionic neuronal somata located in the intermediolateral cell column project through ventral roots to either paravertebral chain ganglia or prevertebral ganglia, as illustrated for the splanchnic nerve (A). Visceral sensory neurons located in the dorsal root ganglia transmit information from innervated visceral organs to interneurons in the spinal cord to complete autonomic reflex arcs at the spinal level (B). *From Loewy and Spyer (1990).*

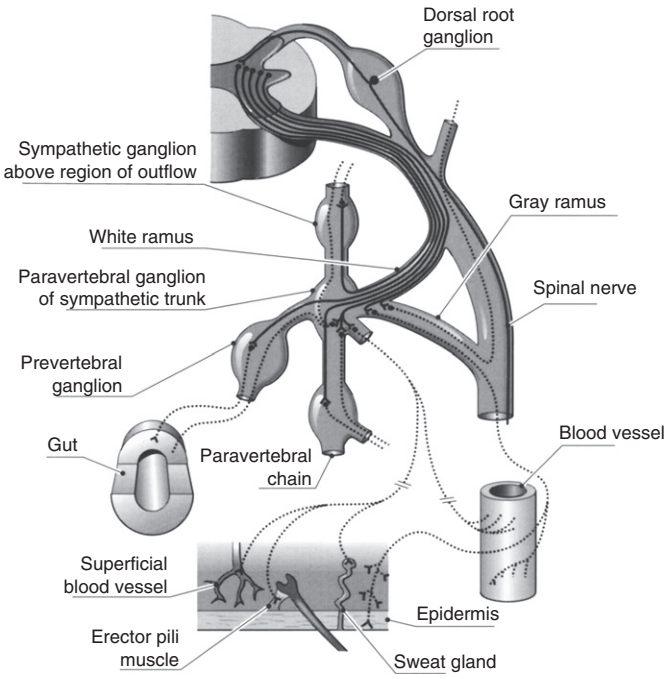


FIGURE 34.3 The SNS is organized segmentally. As illustrated here for one segment of the thoracic spinal cord, preganglionic axons exit through the ventral root of the segment in which the preganglionic cell body is located. These preganglionic axons can be myelinated or unmyelinated and can project to the paravertebral ganglion associated with the segment, to neighboring ganglia of the paravertebral chain, to postganglionic neurons located distally in the prevertebral ganglia, or through one of the autonomic nerves. *From Brading (1999).*

and limbs. For the trunk and limbs, postganglionic axons course both alongside and within the somatic peripheral nerve innervating the region. These autonomic fibers innervate the blood vessels in the muscles (vasoconstrictor fibers), as well as different targets within the skin, including the sweat glands (sudomotor fibers) and erector pili muscles of erectile hairs (pilomotor fibers).

At the rostral end of the sympathetic chain, individual segmental ganglia are fused into aggregate ganglionic groupings. The sympathetic chain, in a modification of the paravertebral chain pattern (Figs. 34.4 and 34.6), innervates the head and the thoracic viscera. The most rostral group, the *superior cervical ganglion* (supplied from the first and second thoracic segments), supplies the head and neck. The targets of its projections are extensive, including the eyes, salivary glands, lacrimal glands, blood vessels of the cranial muscles, and even the blood vessels of the brain. The paravertebral ganglia just caudal to the superior cervical ganglion form the *middle cervical ganglion* and, moving caudally, the *stellate ganglion*. These latter ganglia supply the postganglionic sympathetic outflows that innervate the heart, lungs, and bronchi.

Unlike the trunk, limbs, head, heart, and lungs, organs of the abdominal cavity are innervated by postganglionic

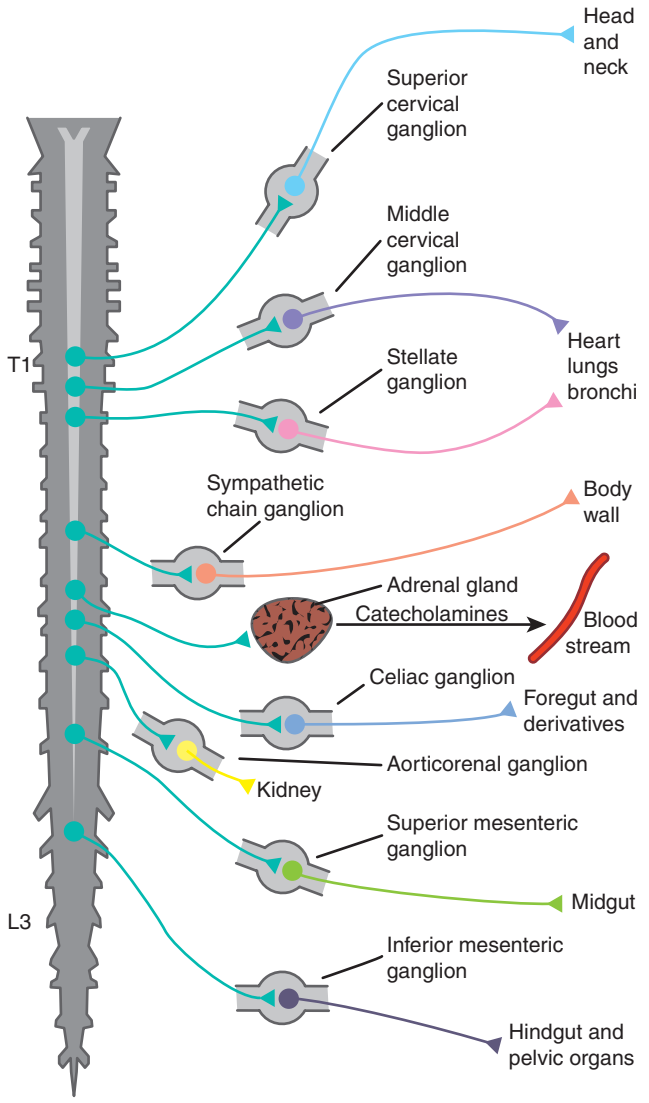


FIGURE 34.4 Summary of the major SNS ganglia and their target organs or tissues. Spinal cord is illustrated on the left. *From Loewy and Spyer (1990).*

neurons in ganglia situated distal to both the spinal cord and the paravertebral sympathetic chain in locations closer to the target tissue. These sites give the ganglia their distinguishing name: *prevertebral ganglia*. The preganglionic neurons innervating these sets of postganglionic neurons are located in the intermediolateral column of the spinal cord, like those innervating the paravertebral chain, but their axons pass through the white rami, the chain ganglia, and the gray rami without synapsing. They then make contact with the peripherally situated prevertebral ganglion cells (Fig. 34.3).

There are four major prevertebral postganglionic stations. The more rostral three are located at points where major abdominal arteries separate from the descending aorta, and the ganglia take their names from their associated arteries (Fig. 34.5). Moving from more rostral to caudal, the *celiac*, *superior mesenteric*, and *inferior mesenteric ganglia* innervate the gastrointestinal tract. The celiac ganglion (also commonly called the solar plexus because its connectives radiate from it) projects predominantly to the stomach and rostral-most parts of the foregut and its embryological derivatives, such as the liver and pancreas. Consistent with their progressively more caudal locations, the middle and inferior mesenteric ganglia innervate the mid- and hindgut tissues, respectively. The inferior mesenteric ganglion also supplies the postganglionic innervation to the pelvic organs. The fourth, and most caudal, prevertebral ganglion is the *pelvic–hypogastric plexus*, which innervates urinary and genital tissues. Perhaps because these prevertebral ganglia are located farther from the spinal cord and closer to the target organs (than the paravertebral ganglia) and because they coordinate responses of visceral organs containing separate nerve networks (the enteric nervous system is described later), these prevertebral ganglia may have more complex neural organization than the paravertebral chain ganglia.

The adrenal medulla is a unique variant of the sympathetic postganglionic pattern. By several criteria, the adrenal medulla is a sympathetic prevertebral ganglion containing postganglionics. However, neurons of this medullary ganglion, rather than issuing axons to innervate target organs, function as a neuroendocrine organ. The adrenal medulla secretes the classical postganglionic sympathetic transmitter norepinephrine, as well as epinephrine, directly into the bloodstream. These catecholamines then circulate as hormones. Release of catecholamines from the adrenals during sympathetic activation provides powerful reinforcement and modulation of the more focal release of catecholamines at traditional—and local—effector sites.

Summary

Overall, the two major motor outflows, or divisions, of the ANS are responsible for the maintenance of homeostasis and support of behavioral responses, but the sympathetic nervous system is the branch specialized for the mobilization of energy. The SNS provides the vascular, glandular, metabolic, and other physiological adjustments that optimize behavioral responses, particularly in emergency situations and conditions requiring activity. The SNS efferent outflow consists of preganglionic neurons, which are located in the intermediolateral column of the thoracic and lumbar spinal cord. SNS efferent axons course to paravertebral and prevertebral ganglia containing the postganglionic neurons, the axons of these postganglionic neurons project to the target tissues.

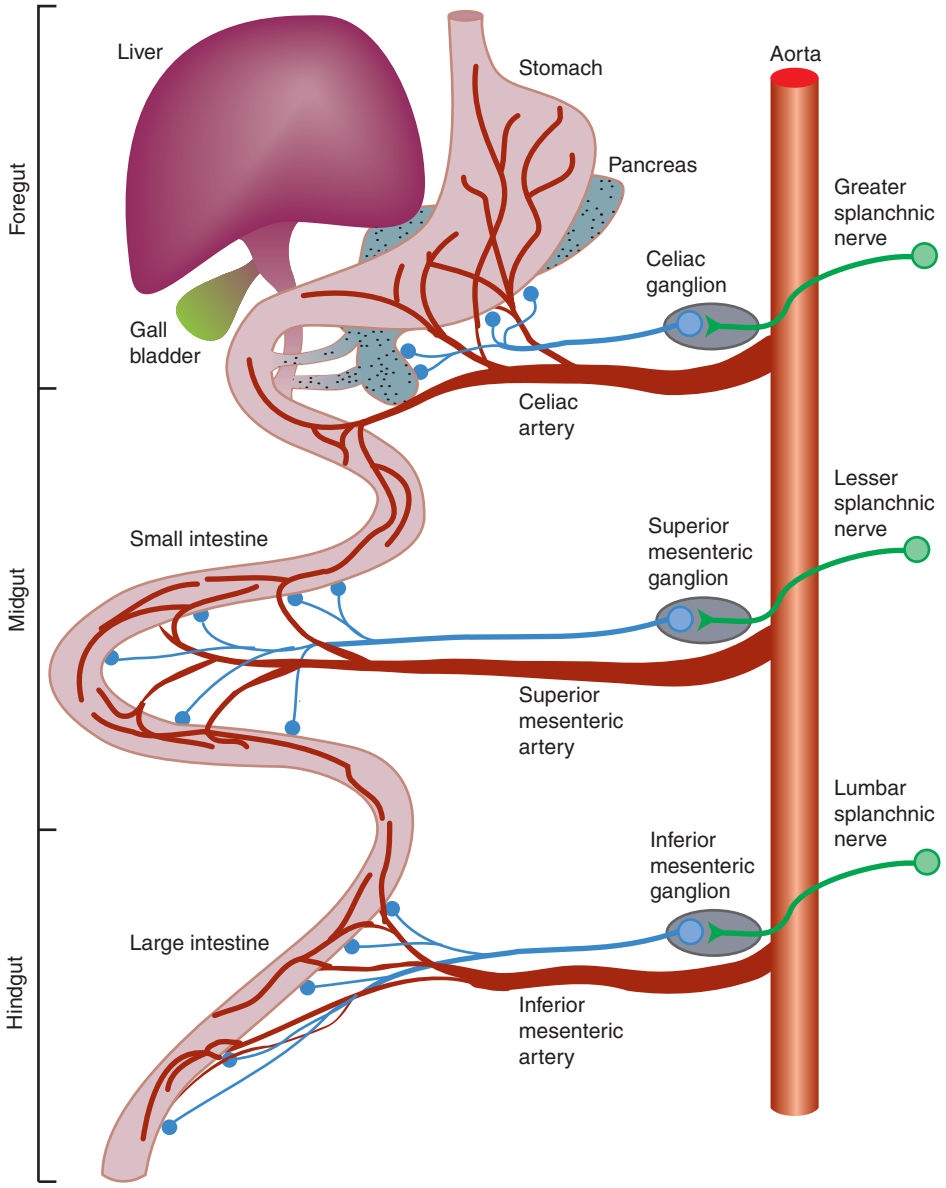


FIGURE 34.5 SNS innervation of the gastrointestinal tract. This more detailed schematic (compared to Fig. 35.4) illustrates how separate segmental levels of the spinal cord and the associated prevertebral ganglia innervate targets in a topographic, specifically viscerotopic, pattern. The association of autonomic pathways and vasculature is also shown. *From Loewy and Spyer (1990).*

PARASYMPATHETIC DIVISION: ORGANIZED FOR ENERGY CONSERVATION

Parasympathetic Nervous System Functions Reduce Energy Expenditure and Increase Energy Stores

Functionally, the parasympathetic nervous system mobilizes homeostatic adjustments that are often opposite and reciprocal to those activated by the SNS (see Table 34.1). In particular, PSNS adjustments usually are considered "rest and digest" responses, in contrast to the "fight-or-flight" sympathetic activation. The PSNS promotes anabolic processing, whereas the sympathetic nervous system augments catabolic activity. The PSNS promotes the gastrointestinal processes required to digest and absorb nutrients effectively. This branch of the ANS also augments the efficient use of energy and the storage of extra calories as fat and glycogen. The PSNS, particularly the vagus nerve, plays a central role in controlling the flow of body reserves or energy between storage depots and the tissues that use the energy. The body's energy reserves are fat in adipose tissue and glycogen in liver and muscle. After a meal, the gastrointestinal tract serves as an additional reservoir of potential energy for the body. By mobilizing triglycerides from adipose tissue or glycogen from liver and muscle, the ANS can make reserves available for metabolism; similarly, by moving nutrient stores (e.g., from the stomach) and promoting their digestion and absorption from the intestines into the bloodstream,

the ANS supplies energy to fuel metabolism or to restock triglyceride and glycogen stores, as needed.

Preganglionic Neurons of the Parasympathetic Nervous System Are Located in the Brainstem and the Sacral Spinal Cord

One of the features distinguishing the PSNS from the SNS is the location of their respective preganglionic neurons. PSNS preganglionic neurons are located in two longitudinal columns of neurons: one in the brainstem and the other in the sacral spinal cord. These locations of PSNS preganglionic neurons are responsible for the alternative name of this division of the ANS: the *craniosacral* (or *bulbosacral*) *division*. The longitudinal column of motor neurons in the brainstem is the general visceral cell column, which condenses into a number of nuclei ventrolateral to the cerebral aqueduct system within the brain. These nuclei then project through cranial nerves to targets in the head (cranial nerves III, VII, and IX) or thorax and abdomen (cranial nerve X) (Fig. 34.6). Cranial nerve III, the oculomotor nerve, carries the axons of the accessory, or autonomic, nucleus of the Edinger–Westphal complex, which participate in the control of the pupillary sphincter and ciliary muscles. Cranial nerve VII, the facial nerve, carries preganglionic axons of the superior salivatory nucleus and controls the submaxillary and sublingual salivary glands, as well as the lacrimal glands. Cranial nerve IX, the glossopharyngeal nerve, carries axons of the inferior salivatory nucleus, which control the parotid salivary glands and mucus secretion.

Preganglionic motor neurons of the Xth cranial nerve, the vagus nerve, control smooth muscles and glands throughout the digestive tract, from the pharynx to the distal colon, including viscera such as the liver and pancreas. These Xth nerve neurons occupy the dorsal motor nucleus of the vagus, a long fusiform nucleus in the medulla oblongata (Fig. 34.7). They control numerous motor and secretomotor responses that participate in the ingestion and digestion of food, as well as in the assimilation of the nutrients. The vagus nerve also carries motor fibers of a special visceral nucleus, the nucleus ambiguus or ventral vagal nucleus, which controls the striated muscles of the pharynx, larynx, and esophagus and the cardiac muscle of the heart. Much like the case described for sympathetic neurons, motor neurons of

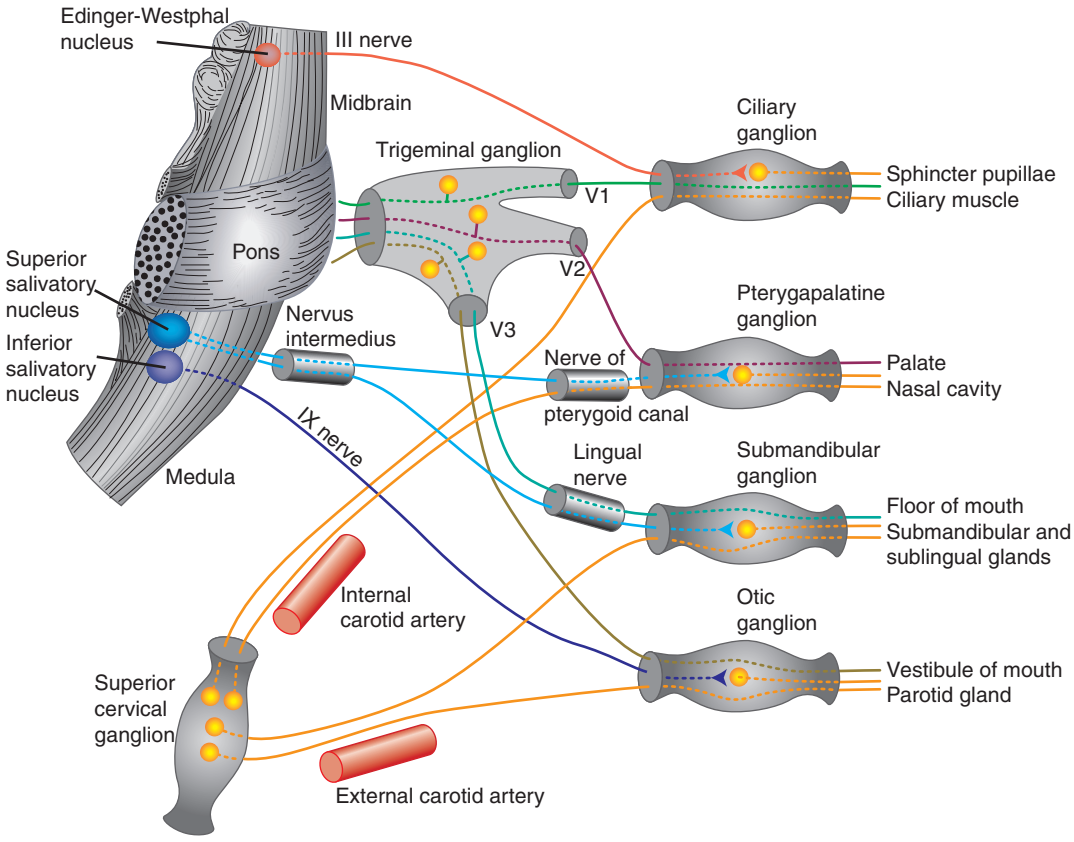


FIGURE 34.6 Autonomic innervation of the head by both sympathetic (superior cervical ganglion) and parasympathetic (cranial nerve nuclei in the medulla and midbrain projecting to the several prevertebral ganglia) pathways. Visceral afferents associated with the autonomic nervous system are found in the trigeminal nerve (cranial nerve V) ganglion. *From Loewy and Spyer (1990).*

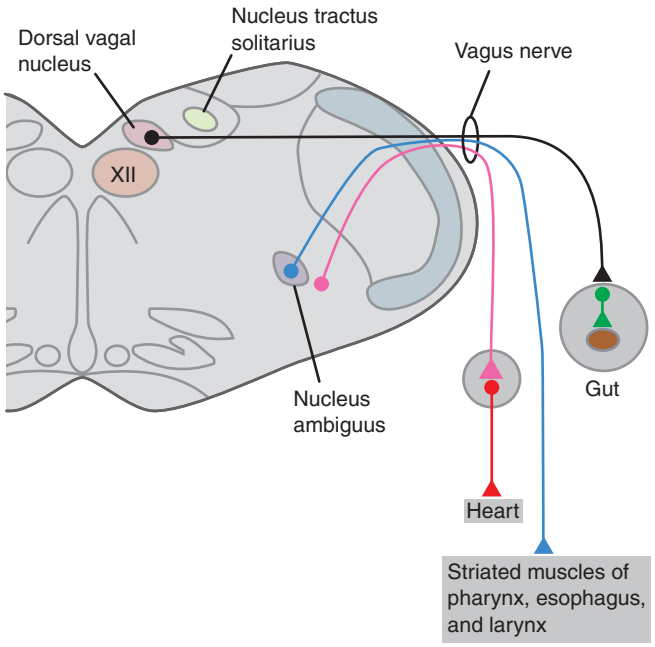


FIGURE 34.7 Parasympathetic projections of the vagal motor pathways, the most extensive of the PSNS circuits, to the viscera of the thorax and abdomen. *From Loewy and Spyer (1990).*

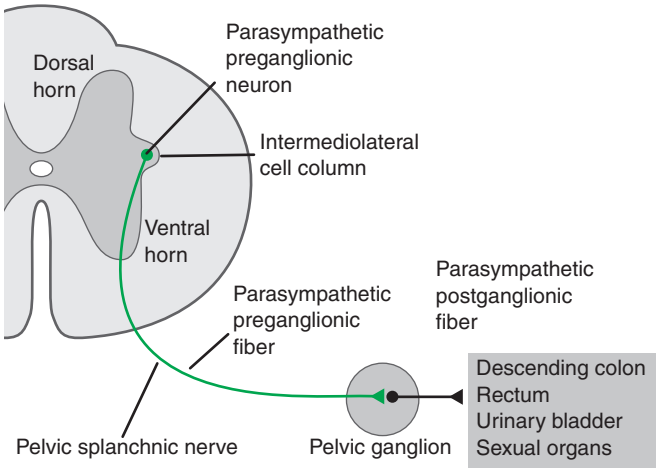


FIGURE 34.8 Spinal cord organization of the sacral division of the craniosacral, or parasympathetic, nervous system. *From Loewy and Spyer (1990).*

the dorsal motor nucleus of the vagus and the nucleus ambiguus are organized into viscerotopic subnuclei (e.g., Bieger & Hopkins, 1987; Fox & Powley, 1992).

The central representation of the more caudal part of the PSNS is found in the sacral spinal cord, in humans the second (S2) through fourth (S4) sacral segments (Fig. 34.8). The parasympathetic preganglionic somata occupy two longitudinal groupings: one column immediately dorsolateral to the central canal and the other in the same transverse location as that of the intermediolateral column of cells, which sympathetic preganglionics occupy in the thoracic and lumbar cord. Axons of the sacral preganglionics exit the cord in the ventral roots, run in a sacral plexus, and project to the target organs. The sacral preganglionic cell columns control parasympathetic motor, vasomotor, and secretomotor functions of the kidneys, bladder, transverse and distal colon, and reproductive organs.

The locations of their respective postganglionic neurons constitute another feature that distinguishes the PSNS from the SNS. Ganglia containing the PSNS postganglionic neurons are juxtaposed to, on the surface of, or even in, the target organ, in contrast to the para- and prevertebral locations of the SNS (Fig. 34.1c). These juxta- and intramural stations often are called *plexuses*, rather than ganglia. Like SNS prevertebral ganglia, the PSNS postganglionic plexuses are complex integrative sites that typically receive inputs not only from preganglionic neurons, but also from afferents within the plexuses or target tissues. Because the PSNS postganglionic neurons are situated so near their targets, the parasympathetic division of the ANS has, perforce, longer preganglionic axons and shorter postganglionic fibers than the sympathetic division. The relative lengths of the postganglionic fibers in the two divisions have been taken as an indication that individual axons of the SNS may diverge more extensively and innervate larger projection fields, whereas the PSNS may exhibit less divergence and more localized projections. (For counterarguments, see Wang, Holst, & Powley, 1995).

Summary

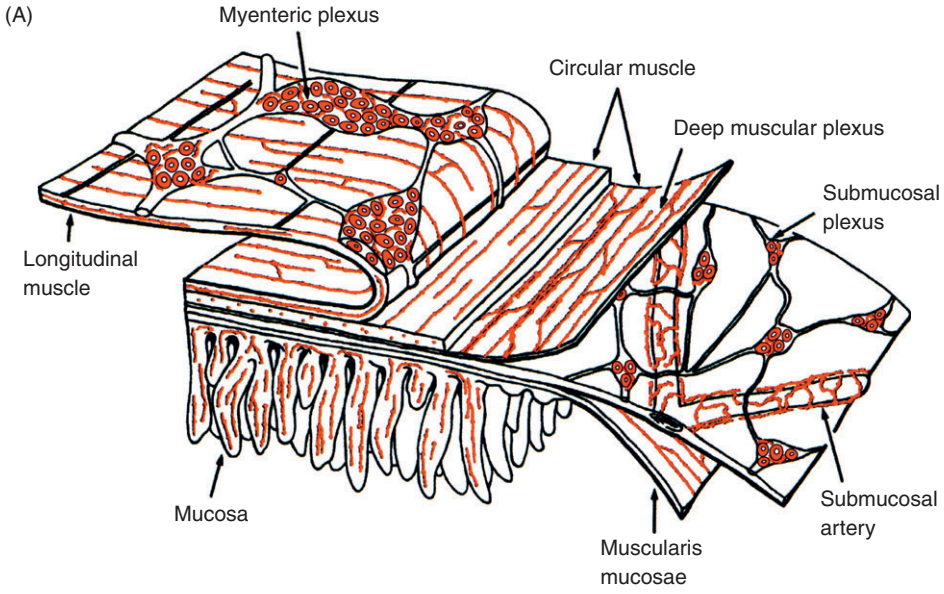
The parasympathetic nervous system or craniosacral division (i.e., the second major outflow of the ANS) is organized to digest, assimilate, and conserve energy, as well as promote rest. Its anabolic functions include not only those associated with the metabolism of nutrients, but also numerous protective reflexes, such as those that limit heat loss, reduce energy expenditure, and slow the heart. The PSNS outflow consists of preganglionic neurons located in brainstem cranial nerve nuclei or the autonomic columns of the sacral spinal cord. Their axons project to the postganglionic neurons located in ganglia situated near or in the target organs. Postganglionic neurons, in turn, project to the smooth muscle of the target viscera.

THE ENTERIC DIVISION OF THE ANS: THE NERVE NET FOUND IN THE WALLS OF VISCERAL ORGANS

Discussions of the ANS sometimes focus on the SNS and PSNS, but there is a third division as well. When Langley articulated the classic definition of the ANS, he

identified the *enteric nervous system* (ENS) as the third division. The alimentary canal, or "entrum," and the tissues derived from it, such as those in the pancreas and liver, contain extensive and well-formed neural networks. In particular, the gastrointestinal tract contains two major plexuses that contain as many neurons as the entire spinal cord (Fig. 34.9). Each plexus consists of an extensive sheet of small nodes of cells linked by connectives. One plexus, the myenteric plexus, is situated between outer longitudinal and inner circular muscle layers of the viscus. The other plexus, the submucosal or submucous plexus, is located between the circular muscle layer and the lumenal mucosal layer of the viscera. As illustrated in Figure 34.9, there are additional, less extensive, enteric networks as well.

The number of neurons in the enteric plexuses, their extensive interconnections, and their capacity to support motility have led to proposals that the enteric nervous system serves as a more or less independent "brain" in the gut (Wood, 1987). Early descriptions emphasized



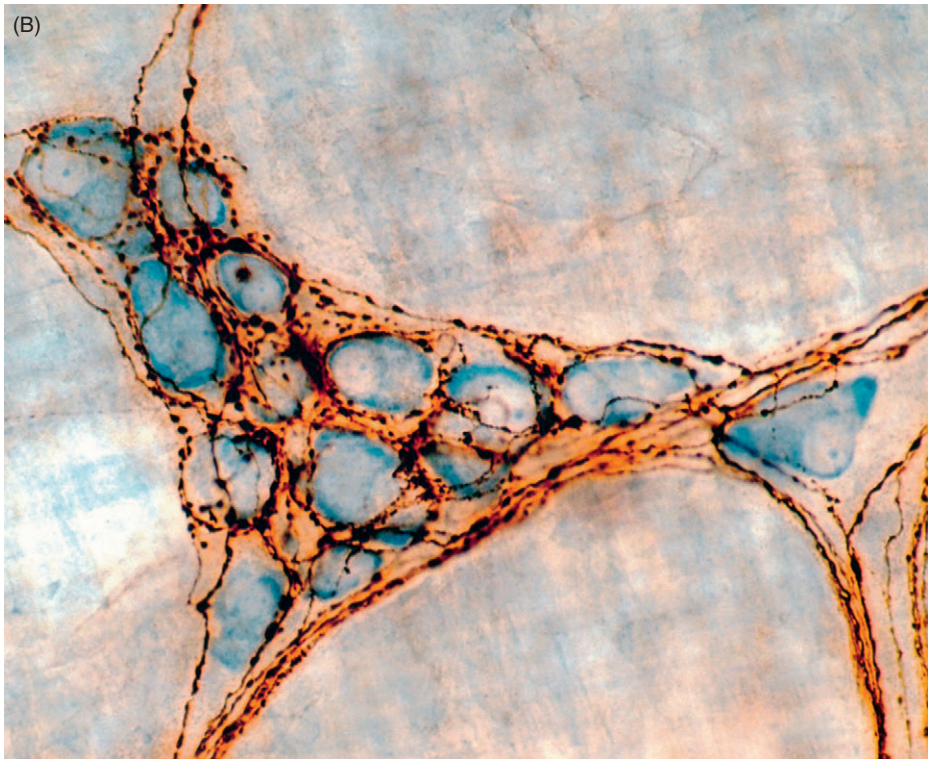


FIGURE 34.9 (A) The enteric nervous system plexuses in the wall of the intestine. Enteric neurons form two conspicuous and extensive networks, or plexuses, of ganglia and connectives located between the longitudinal and circular muscle layers of the wall (the myenteric plexus) and the circular muscle layer and the inner mucosal layer (the submucosal plexus). The deep muscular plexus, periglandular plexus, and villous plexus are additional networks of enteric neurons and their processes. *From Furness (2006).* (B) Autonomic preganglionic terminals innervating postganglionic neurons. This example of vagal preganglionic projections (brown, labeled with the tracer *Phaseolus vulgaris*) innervating myenteric ganglion neurons (blue, stained with Cuprolinic blue) in the stomach wall illustrates that autonomic preganglionic axons can form extensive networks controlling postganglionics. *From Holst, Kelly, and Powley (1997).*

the autonomy of the ENS by observing that the gastrointestinal tract could exhibit movement, peristalsis, even after all extrinsic connections to it were cut. However, fully integrated or coordinated responses of the gastrointestinal tract, including motor responses other than peristalsis, as well as absorptive and secretory responses, involve extrinsic projections from the central nervous system to the enteric nervous system.

Because the preganglionic neurons of the vagus nerve project to the enteric ganglia rather than to independent postganglionic stations, the ENS serves as an extensive postganglionic relay station for the vagus (Fig. 34.9B).

Summary

The ENS or enteric nervous system, the third division of the ANS, consists of ganglia and plexuses of intrinsic neurons in the walls of the viscera such as the gastrointestinal tract. These nerve networks contain intrinsic afferents, interneurons, and efferents that control local functions; they also receive extrinsic inputs from autonomic preganglionic efferents, particularly those PSNS projections coursing in the vagus nerve.

ANS PHARMACOLOGY: TRANSMITTER AND RECEPTOR CODING

Early in the twentieth century, pharmacological studies performed with naturally occurring agonists and antagonists provided the basis for the original differentiation of two major ANS divisions: the SNS and PSNS. Prior to the pharmacological analyses, the two often had been lumped together. Dale (the contributor of "Dale's law" of neurotransmitters), Langley, and a number of other early investigators applied natural extracts such as muscarine, nicotine, *d*-tubocurarine, and others to autonomic ganglia while measuring responses in order to infer the chemical taxonomy of the ANS. Subsequent experiments have replicated the basic distinctions made by these investigators, but this newer research, with access to a pharmacopoeia of synthetic agonists and antagonists, as well as to immunocytochemistry for the characterization of transmitter substances, has described a much richer and more complicated multiplicity of neurotransmitters, neuromodulators, and receptor subtypes in autonomic nerves (Lundberg, 1996).

Preganglionic Neurons Use Acetylcholine as a Transmitter

Most preganglionic neurons of both the SNS and PSNS have a cholinergic transmitter phenotype. Only relatively small subpopulations of preganglionic neurons express other (e.g., dopaminergic or adrenergic) phenotypes. In the peripheral ganglia of these two divisions of the ANS, most postganglionic neurons, which receive inputs from the preganglionics, express a predominance of a nicotinic form of the cholinergic receptor on their somata. As discovered by the early autonomic pharmacologists, nicotine can serve as a general "ganglionic blocker" to thwart transmission at these synapses. Additional specificity and response selectivity, or coding, at the ganglionic level normally are maintained by the segregation of sympathetic and parasympathetic postganglionic neurons in separate ganglia, by point-to-point axonal projections, by distinguishing neuropeptides colocalized with the transmitter, and by the spatial separation of synapses within ganglia.

SNS and PSNS Postganglionic Neurons Use Different Transmitters

The respective transmitter phenotypes of the postganglionic neurons constitute one of the cardinal defining differences between the sympathetic and parasympathetic nervous system. The majority of SNS postganglionic neurons release the transmitter norepinephrine, whereas PSNS postganglionic neurons release primarily acetylcholine. Such a two-transmitter chemical code produces push–pull or positive bidirectional control of individual targets.

The biomechanics of smooth muscle are quite different from the biomechanics of striated muscle. In contrast to skeletal muscle, smooth muscle is not always organized into antagonistic pairs, with each member of the pair being innervated by a different motor neuron. Also, the smooth (and cardiac) muscles forming the walls of the viscera are not differentially attached to a hard frame to provide mechanical leverage through opposal activity. Reciprocal motor programs of the viscera, for the most part, involve active, phased contraction and relaxation of the same muscle to achieve coordination. To produce these patterns of excitation and inhibition, the target organs and tissues express both adrenergic and cholinergic receptors and have different, often antagonistic or complementary, responses to selective activation of the different receptors.

Response patterns of smooth muscle are further differentiated by specializations of the receptors. Adrenergic receptors, influenced by the catecholaminergic transmitter released by sympathetic fibers, are differentiated into at least four different metabotropic subtypes linked to different intracellular pathways. Similarly, the muscarinic acetylcholine receptor, influenced by the cholinergic transmitter released by parasympathetic postganglionic neurons, is differentiated into at least three ionotropic subtypes with differential influences on intracellular transduction. These heterogeneities of transmitter species and receptor types make it practical to mobilize multiple, different, and

BOX 34.1

AUTONOMIC POSTGANGLIONIC NEURONS CAN CHANGE THEIR TRANSMITTER PHENOTYPES

Transmitter phenotypes of autonomic neurons generally seem fixed, dictated by both intrinsic and extrinsic developmental cues. At least some ANS neurons, however, can alter their phenotypes under appropriate environmental conditions. This conclusion first was reached when populations of postmitotic adrenergic and cholinergic neurons were cocultured *in vitro* in different media or in the presence of different tissues. Without changes in overall cell number, the percentages of adrenergic and cholinergic neurons in a given culture changed over time. Even more definitively, when single postganglionic cells were grown in microculture wells, the presence of a medium conditioned by heart cells caused some of these mature neurons to change from an adrenergic to a cholinergic phenotype (Potter, Landis, & Furshpan, 1980).

This phenotypic switching does not appear to be an artifact of cell culture situations; the change appears to occur normally *in vivo* in development to generate the specialized minority of sympathetic postganglionic neurons that are cholinergic (e.g., for innervation of sweat glands). Landis and Keefe (1983) have verified this hypothesis by using rat foot pad sweat gland innervation *in vivo* and characterizing the transmitter phenotype of the autonomic projections at different stages of development. The developing sympathetic postganglionic innervation of the sweat glands initially expresses a typical adrenergic phenotype as the axon grows toward its target. Once the sweat gland target is innervated, however, the tissue induces the change in neurotransmitter phenotype to a cholinergic one.

Such changes in transmitter are not unique to these examples. Perhaps reflecting pluripotential patterns seen earlier in phylogeny, at least some cholinergic neurons (e.g., avian ciliary ganglion cells) appear to remodel into an adrenergic phenotype, whereas other cells (e.g., maturing neural crest cells in the developing gut wall) exhibit a transient catecholaminergic stage that is lost during the ingrowth of extramural innervation.

Finally, in addition to such switching, autonomic postganglionic neurons also have been shown to exhibit plasticity in altering their levels of transmitters and cotransmitters as a function of the activity of their preganglionic inputs. For example, after preganglionic inputs are blocked or transected pharmacologically, postganglionic neurons of the superior cervical ganglion decrease their levels of tyrosine hydroxylase while increasing their level of substance P.

Terry L. Powley

potentially highly differentiated responses from one effector tissue (Box 34.1).

Response selection may also be facilitated by neuropeptides that the postganglionic neurons synthesize and corelease during activity. Somatostatin, neuropeptide Y, or both are found in many sympathetic neurons; vasoactive intestinal polypeptide and calcitonin gene-related peptide are often coexpressed in parasympathetic neurons. These peptides may serve as neuromodulators that vary postsynaptic responses to the conventional autonomic neurotransmitters (e.g., Costa, Furness, & Llewellyn-Smith, 1987). The release of different neuropeptides and cotransmitters can vary with different rates or patterns of firing and is not invariantly proportional to transmitter release. Thus, a heterogeneity of different cotransmitter and transmitter release patterns can be produced by a single fiber or fiber type, depending on firing pattern and other local factors at the synapse (e.g., Lundberg, 1996).

Although dual innervation in an oppositional or push–pull pattern is presumed to yield faster, more responsive adjustments and to provide a mechanism that can adjust response gain, some tissues are innervated by only one arm of the ANS. The SNS has a wider distribution than the PSNS. In particular, white and brown adipose tissue, peripheral blood vessels, and sweat glands are innervated by the SNS without reciprocal PSNS projections. One possible explanation for this arrangement is that the functions of these particular effectors do not require the speed associated with push–pull projection patterns, particularly in terminating responses once activated. Furthermore, in the evolution of the autonomic nervous system, the SNS is thought to be the older, and correspondingly more extensive, network, whereas the PSNS occurred more recently in phylogeny and therefore has less widespread projections.

Finally, the enteric nervous system also has an extensively varied set of transmitters, cotransmitters, and neuromodulators, providing highly differentiated chemical coding of local functions (Furness, 2006; Lundberg, 1996).

Summary

Both the SNS and the PSNS use acetylcholine as a preganglionic transmitter, but they are distinguished by different postganglionic phenotypes: most SNS postganglionic neurons are catecholaminergic; most PSNS postganglionics

are cholinergic. The different neurotransmitter and cotransmitter elements expressed in the postganglionic neurons, as well as numerous postsynaptic receptor subtypes, provide the chemical coding for responses of the PSNS and SNS. This functional organization ensures a varied, powerful, and dynamic push–pull operation of homeostatic systems.

AUTONOMIC COORDINATION OF HOMEOSTASIS

The "self-governing" or autonomic characteristic that gives its name to the ANS is based largely on reflexes that involve more than peripheral efferent pathways. Though the early history of autonomic neuroscience emphasized this motor outflow, homeostasis is maintained through the cross-coupling and coordination of multiple reflex arcs. These reflexes involve afferent inputs as well as efferent outputs.

Two major inflows of visceral, or autonomic, afferents (Box 34.2) establish the input arms of the basic reflex arcs. These visceral afferents also provide critical cross-linked feedback between the sympathetic and the parasympathetic outflows, keeping them in balance. For example, sympathetic activity can increase heart rate (tachycardia) and produce constriction of strategic vascular beds, thus leading to an increase in blood pressure. This increase in blood pressure, in turn, is detected by vagus nerve baroreceptors in the aortic arch and other sites. Visceral afferents in the vagus can then reflexively stimulate vagal parasympathetic efferents that slow heart rate (bradycardia). The visceral afferents monitoring the effects of autonomic activity produce positive, reciprocal, and dynamic regulation of physiological responses. When a more sustained adjustment in blood pressure is required (e.g., in the general arousal of the fight-or-flight response), these outflows are adjusted centrally in a process akin to the α–γ linkage so that sympathetic activation is not nullified by parasympathetic responses damping the needed activation.

Visceral Afferents Link Target Tissues and the CNS

In the case of the sympathetic nervous system, motor neurons receive direct afferent input from most of the autonomic targets. The cell bodies of these primary afferents are located in the dorsal root ganglia of the spinal cord segment(s) in which the corresponding preganglionic neurons are located (Figs. 34.2B and 34.10). Afferent somata are similar to other dorsal root ganglion cells, although they frequently are distinguishable as relatively small, dark neurons or B afferents. From endings in the innervated viscera or tissue, the centripetally directed

BOX 34.2

AUTONOMIC REFLEXES: ACTIVATED BY VISCERAL AFFERENTS, AUTONOMIC AFFERENTS, OR JUST AFFERENTS?

A long-standing dispute about terminology still influences discussions of autonomic reflexes. Some authors (and textbooks) adhere to the original autonomic classification of Langley, who considered the ANS a motor system without a corresponding sensory inflow. This classic view acknowledges that autonomic effectors are influenced by afferents but considers visceral afferents, as well as somatic afferents, as independent of the ANS. Conversely, other scientists (and textbooks) consider afferents arising in and associated with the viscera as the necessary counterpart to autonomic efferents and label them autonomic afferents.

The controversy goes back to Langley's concentration on a pharmacological definition of the ANS. In his initial studies, Langley speculated about the existence of autonomic afferents, but was unable to find a neurochemical marker to distinguish visceral from somatic afferents. He sidelined the issue pending more information (his seminal book is subtitled Part I), focused his studies on efferents, and never returned to the search for an afferent marker (Part II never appeared). Adherents to his original motor-only classification stress that somatic and visceral afferents can elicit autonomic responses and that both classes of afferents elicit somatic responses, as well.

However, advocates for including visceral afferents within the autonomic schema argue that these afferents innervate the target tissues of the ANS efferents, share similar embryological histories with autonomic efferents, course in the same peripheral nerves as the autonomic efferents (e.g., the vagus), and form mono- and oligosynaptic reflex arcs influencing autonomic preganglionic neurons. These proponents for revision of the autonomic terminology also point out that more recent analyses have identified neuropeptide markers shared by visceral afferents (but not somatic afferents) and autonomic motor neurons. This more inclusive view of the autonomic nervous system seems to be gaining ground. It works more naturally for the discussion of reflexes, and we have adopted it in this text.

Terry L. Powley

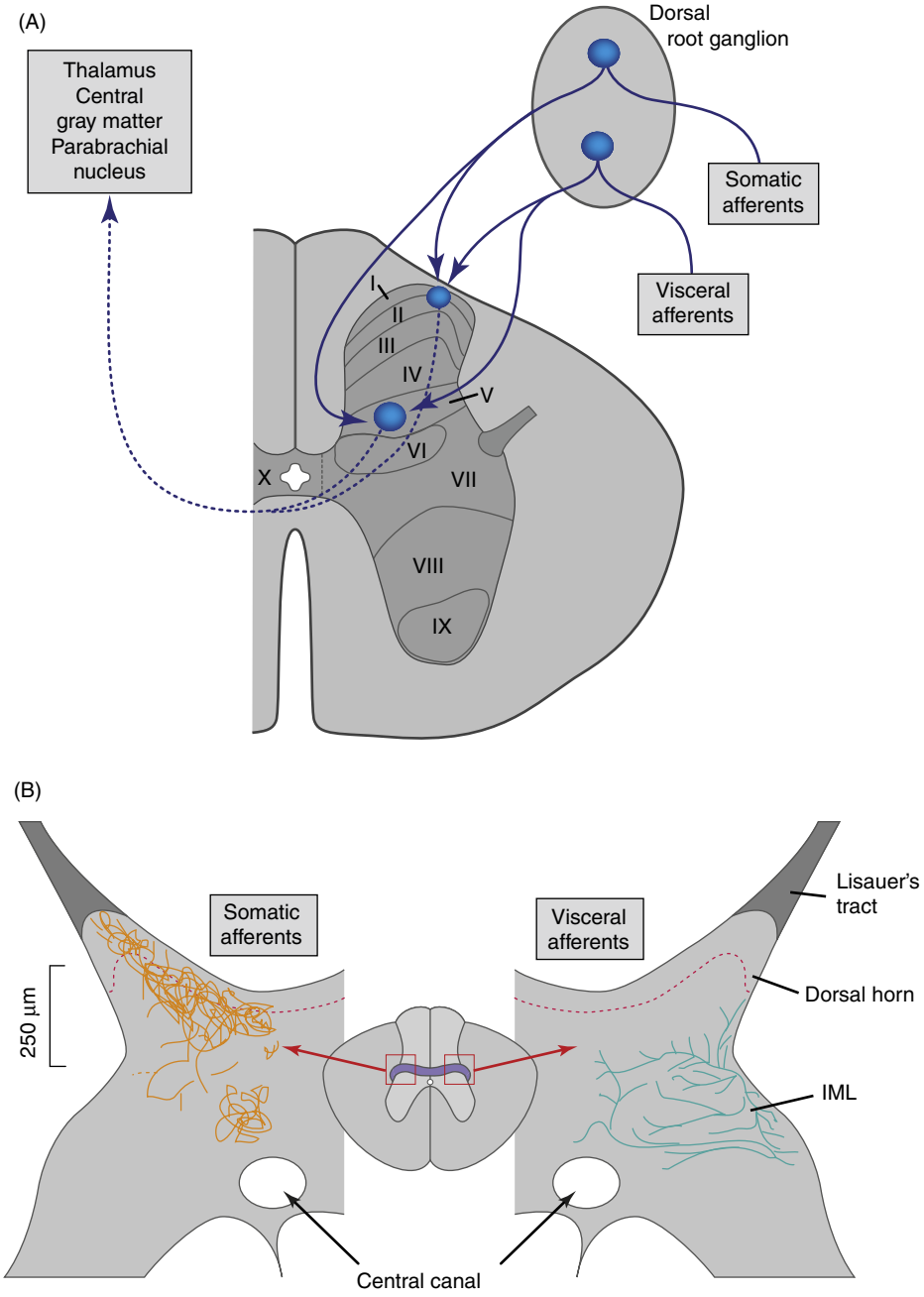


FIGURE 34.10 Visceral and somatic afferents follow parallel but distinct paths in the central nervous system. Visceral and somatic afferents consist of different populations of dorsal root ganglion neurons that project to laminae I and V of the dorsal horn of the spinal cord. These relay sites provide local spinal reflexes and also project to higher autonomic and somatic sites, respectively, in the brain (A). Although visceral and somatic afferents follow similar trajectories, more detailed analyses indicate the two types of afferents end in distinctly different distributions and densities within the spinal cord (B). IML, intermediolateral cell column. *From Cervero and Foreman (1990).*

peripheral processes of the afferents typically course in mixed nerves that also contain the motor outflow. These processes then reach the dorsal roots through the major peripheral nerves. The central processes of these visceral afferents enter the spinal cord in association with Lissauer's tract, ending in laminae I and V of the dorsal horn (Cervero & Foreman, 1990). For many of these visceral afferents, their endings in the periphery and in the spinal cord contain substance P (SP) and other neuropeptides of the tachykinin family, such as neurokinin A (NKA) and neurokinin B (NKB) (see Lundberg, 1996).

Visceral afferent nerves relay sensory information about visceral volume, pressure, contents, or nocioceptive stimuli to spinal centers, where automatic responses are interpreted and reflex responses are generated.

In the parasympathetic limb of the ANS, there are two major afferent inflows with different organization. In the cranial division, visceral afferents most often are found in the same cranial nerve as their corresponding motor counterparts. These mixed cranial nerves have sensory ganglia located outside the cranium, and the cell bodies of the afferents are found in these ganglia. Like other visceral afferents, these neurons are typically pseudounipolar neurons with a peripheral neurite extending to the target tissue and a central process projecting to the CNS. The central terminals of these afferents end in a cranial nerve sensory

nucleus. Visceral afferents associated with the outflow of cranial nerve III are complex and enter through different channels (e.g., cranial nerve V, the trigeminal nerve). Most of the visceral afferents associated with cranial nerve autonomic reflexes terminate in the extensive nucleus of the solitary tract, which is located immediately dorsal to the general visceral column of the brainstem. Specifically, afferents of VII, IX, and X all end in the nucleus of the solitary tract. These inputs form a viscerotopy in the nucleus, with facial gustatory information projecting most rostrally and medially, the glossopharyngeal information somewhat more caudally, and the afferents of the different branches of the vagus nerve terminating most caudally in different subnuclei of the nucleus of the solitary tract. As the largest and most complex of the visceral afferent relays in the brain, the nucleus of the solitary tract also receives afferent inputs from the spinal division of the trigeminal nerve and second-order inputs from the dorsal column nuclei.

Like visceral afferents associated with the SNS, visceral afferents associated with the PSNS contain SP and other tachykinins. In addition, several neuropeptides found in the gut are also found in some of these afferents. For example, the gut hormone and neuropeptide cholecystokinin (CCK) is found in vagal afferents relaying information from the gastrointestinal tract to the brain. CCK is also found in higher-order ascending relays of vagal projections in the neuraxis, leading some to argue that CCK provides chemical coding for visceral afferents associated with the gastrointestinal tract.

The second major inflow of afferents to the parasympathetic limb of the ANS is associated with the sacral division. Visceral afferents in this division are organized much like the spinal afferents associated with the thoracolumbar or sympathetic division of the ANS (e.g., Morgan, Nadelhaft, & de Groat, 1981).

At the central relays of autonomic reflex arcs, the thoracolumbar spinal circuitry of sympathetic reflexes and the craniosacral circuitry of parasympathetic responses have similar morphological elements. The SNS motor neurons preferentially distribute their dendrites within the long longitudinally oriented column that constitutes the intermediolateral nucleus of the spinal cord. The neurons also distribute a subset of their dendrites dorsolaterally in an arch that brings the dendrites into contact with the incoming visceral afferents (e.g., Dembowsky, Czachurski, & Seller, 1985). Parasympathetic preganglionics exhibit similar dendroarchitecture (e.g., Fox & Powley, 1992). This characteristic dendroarchitecture is consistent with the ideas that local mono- or oligosynaptic reflexes are organized segmentally and at the preganglionic outflows.

Visceral Afferents Also Organize Axon Reflexes and Signal Visceral Pain

In addition to forming conventional reflex circuits throughout the ANS, at least some visceral afferents support effector responses known as *axon reflexes*. Unlike a conventional reflex, which typically involves an afferent neuron, at least one central nervous system synapse, and an efferent neuron, axon reflexes involve only the afferent neuron and they occur in the target tissue—without a CNS relay. The phenomenon can be seen clearly in the type of experiment that was instrumental in establishing the existence of axon reflexes. If all motor axons projecting to a target tissue are eliminated by surgery or other appropriate means while the afferent innervation of the site is spared, the afferent axons can be stimulated selectively. When the afferent axons are stimulated such that their peripheral processes innervating the target are invaded antidromically, one can measure an effector response or assay the release of a neuropeptide from the peripheral afferent neurite. Such experiments performed on a variety of different visceral and cutaneous afferent systems have demonstrated that axon reflexes produce a number of inflammatory and vascular effector responses (Lundberg, 1996).

When a peripheral ending of a visceral afferent transduces a stimulus, the resulting action potential can release the neuropeptide contained in vesicles in that ending (Fig. 34.11). The released compound then acts as

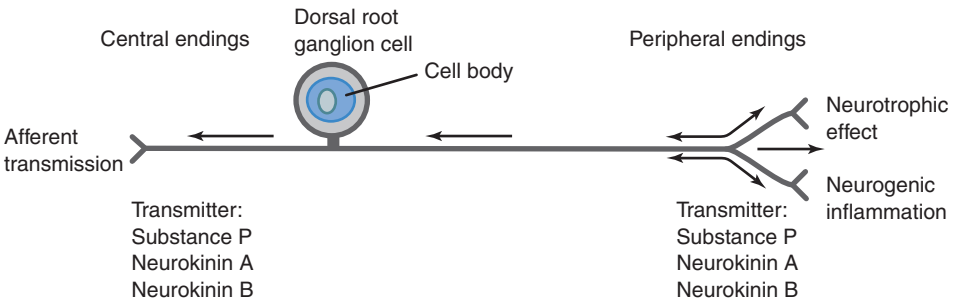


FIGURE 34.11 The architecture of visceral afferents that produce axon reflexes. Visceral and cutaneous afferents release transmitters from the tachykinin family. When action potentials (indicated by arrows) are generated peripherally, they are propagated centrally. Peripherally generated action potentials also produce a local release of tachykinins in the affected terminals and in terminals of the peripheral collaterals. *From Cervero and Foreman (1990).*

a neuromodulator or neurotransmitter to produce local changes on smooth muscle—a local, or peripheral, axon reflex. The response is also relayed to neighboring tissues. When an action potential is transmitted centrally in an afferent axon, it can also be transmitted centrifugally by collaterals of the same fiber, and release of the neuromodulator from these collateral endings can propagate the reflex in the immediate area. The inflammatory responses and extravasation associated with the classic "wheal and flare" reaction to skin damage were the first axon reflexes analyzed and are a classic illustration of the phenomenon. Such local responses have been widely documented not only in the skin, but also in viscera such as the lungs.

Finally, visceral afferents, particularly those associated with SNS pathways, are responsible for visceral pain (Craig, 2002; Jänig, 1996).

Summary

Visceral afferents innervate the target tissues of the ANS, and their axons frequently course in the same peripheral nerves as autonomic efferents. Also commonly called autonomic afferents, these sensory elements associated with the ANS constitute the afferent limbs of autonomic reflexes and carry the inputs recognized as visceral pain. Visceral afferents also mediate axon reflexes and are cross-linked with somatic, as well as autonomic, nervous system activity.

HIERARCHICALLY ORGANIZED ANS CIRCUITS IN THE CNS

Homeostasis no more occurs through isolated autonomic reflexes than posture is maintained, or movement is executed, through isolated somatic reflexes. Nevertheless, methodological and historical considerations have yielded a tendency to focus on individual autonomic reflexes. Methodologically, the need for experimental control typically dictates that experiments isolate a reflex (or small set of reflexes) for examination. In addition, historically, the ANS was conceptualized primarily as a motor branch of the peripheral nervous system.

Though the separate-reflex perspective has in some ways shaped a view of ANS functions as a collection of individual motor responses, autonomic activity typically involves finely coordinated, fully integrated adjustments of multiple outflows. Just as somatic posture and movement are coordinated by supraspinal CNS controls, autonomic integration is achieved by supraspinal controls. The preganglionic motor pools of the craniosacral and thoracolumbar outflows are final common paths for descending projections from the CNS. Similarly, the first- and second-order visceral afferent relays (e.g., the nucleus of the solitary tract and the spinal lamina V), which form short, mono-, or oligosynaptic reflex arcs with these preganglionic motor pools, are targets of rostral CNS stations that coordinate and modulate the autonomic outflows.

The CNS coordination of autonomic activity achieves a number of integrative functions:

- Much like the suprasegmental organization of skeletal motor responses, central ANS stations provide coordination and sequencing of different local autonomic reflexes. These stations provide, for example, the efferent choreography necessary to coordinate the sequential autonomic responses in the mouth, stomach, intestines, and pancreas during and after a meal. Such programs coordinate brainstem and spinal cord efferents, as well as reflexes within the SNS and PSNS divisions of the autonomic outflow.
- Central ANS circuitry also cross-links autonomic activity and somatic motor activity. The example used earlier of cardiovascular adjustments occurring in concert with postural adjustments illustrates such linkage between autonomic and somatic nervous system outflows.
- The CNS stations of the autonomic nervous system also provide the information, as well as the organization and planning, that enables an organism to mobilize autonomic adjustments or responses in anticipation of environmental contingencies and developing needs. For example, both fluid and energy homeostasis involve physiological adjustments in anticipation of major imbalances or deficits, responses that cannot be explained solely by reactive reflexes.

The importance and the nature of central autonomic controls are underscored by clinical disorders that result from interruptions in the connections between the central autonomic circuitry and the preganglionic motor neurons (Box 34.3). Quadriplegia and paraplegia, which result from injuries that divide the separate thoracic and lumbar sympathetic loops and sacral parasympathetic pathways from higher central regions, illustrate how essential these longitudinal connections are. Interruption of the longitudinal connections causes loss of voluntary bladder and bowel control. Disruption of the spinal cord also causes men to become impotent because penile tumescence is a predominantly parasympathetic response and ejaculation is a sympathetic response. Although erections can occasionally be elicited as local reflexes caused by direct mechanical stimulation, ejaculations, which normally require suprasegmental coordination, almost universally disappear. In addition, numerous vascular and glandular reflexes are also disordered by the interruption of long autonomic pathways connecting the brain and spinal cord (Bannister, 1989).

BOX 34.3

DISEASES AND AGING IMPAIR THE AUTONOMIC WISDOM OF THE BODY

The importance of the autonomic nervous system to health often goes unnoticed. When ANS functions operate normally, they typically occur automatically, without the individual being aware of any adjustments. Nevertheless, loss of an autonomic response can be disruptive, and autonomic disorders can be debilitating (Bannister, 1989). In a classic example of ANS degeneration known as multiple system atrophy and autonomic failure, or the Shy–Drager syndrome, individuals exhibit postural hypotension, urinary and fecal incontinence, sexual impotence, loss of sweating, cranial nerve palsies, and a movement disorder similar to Parkinson's disease. Disconnection of the suprasegmental autonomic stations from the spinal cord can occur in neurological disorders, such as multiple sclerosis, and traumatic spinal cord injury. As a result, bowel and bladder control are lost, and impotence is caused by a loss of autonomic genital reflexes.

Some autonomic disorders are less conspicuous. In part, the apparently more subtle deficits may be ones that are less important in the carefully controlled environments found in modern societies. For example, impaired thermoregulatory responses may be tolerated well by a person living in a climate-controlled environment. Also, metabolic emergencies are unlikely to occur and need autonomic compensation if a person regularly eats enough food. Nevertheless, autonomic disorders occasion a number of serious, even life-threatening, health problems.

Autonomic imbalances can take many different forms. Excess activation of the ANS is implicated in various stress-related disorders, including ulcers, colitis, high blood pressure, and heart attacks. As discussed earlier, chronic failure of autonomic cardiovascular homeostasis can cause debilitating postural hypotension. Developmental autonomic disorders, such as immature or anomalous medullary respiratory circuitry, are thought to contribute to sudden infant death syndrome. Some sleep disorders, such as sleep apnea, are also thought to have autonomic elements. Furthermore, widespread autonomic dysfunction resulting from autonomic neuropathies of diabetes, alcoholism, and Parkinson's disease complicates the primary diseases. Autonomic disturbances can also underlie metabolic disorders, such as stress-induced diabetes and reactive hypoglycemia. Autonomic disturbances have also been hypothesized to be a cause of obesity.

Autonomic function is not only impaired by diseases, it is also compromised by aging processes—even those of healthy aging (Amenta, 1993). With aging, both cell losses and neuropathies of surviving neurons occur in ANS circuits. Such age-related changes in the ANS reduce an individual's ability to optimize and sustain homeostatic responses. These consequences of aging produce both primary autonomic disorders and secondary complications and vulnerabilities when the individual is challenged by disease or injury.

Terry L. Powley

Early Stimulation and Lesion Experiments Identified a Central Component of the ANS

Like the concept of the ANS itself, the corollary that the ANS includes a hierarchy of central mechanisms was first demonstrated by physiological experiments. Nineteenth- and early twentieth-century experiments manipulating the brain caused autonomic disturbances. In one classical experiment, for example, Claude Bernard demonstrated that localized mechanical lesions, or stab wounds, in the floor of the fourth ventricle near the dorsal vagal complex produced a "piqure glycosurique," a disturbance of blood glucose homeostasis characterized by a diabetes-like condition in which glucose spills into the urine. Subsequent lesion and stimulation studies identified a number of "centers" for micturition, respiration, and other autonomic functions in the medulla oblongata and pons.

Other experiments implicated the hypothalamus in even more extensively coordinated and cross-linked autonomic patterns. This work led to the concept of the hypothalamus as the "head ganglion" of the autonomic nervous system. Following an earlier application of electrical stimulation to the hypothalamus by Karplus and Kreidl, Walter Hess found that focal electrical stimulation of regions of the hypothalamus elicited coordinated patterns of sympathetic and parasympathetic adjustments (e.g., changes in pupil size, piloerection, and respiration) and affective responses that were appropriate to behavioral responses elicited from the same loci. In fact, central manipulations that affect autonomic function seem practically invariably to produce adjustments of both SNS and PSNS outputs, an observation that underscores the integrative role the brain plays in ANS function.

Brain lesions in virtually all limbic system regions, and notably in the septal area, amygdala, hippocampus, frontal cortex, cingulate cortex, and insular cortex, have been shown to exaggerate, dissociate, blunt, or in other ways distort autonomic responses to environmental situations.

Anatomical Mapping Experiments Have Identified Multiple Hierarchically Organized, Reciprocally Interconnected Stations of the Visceral Neuroaxis That Control Autonomic Activity

As neuroscience tracing tools have become more powerful, they have revealed an extensive CNS autonomic control hierarchy coupling visceral afferent inputs with autonomic efferent outflows, linking the SNS and PSNS divisions of the ANS, and interconnecting the central autonomic stations with somatic and endocrine pathways in the brain (Box 34.4). This central autonomic circuitry has been considered a "central visceromotor system" (Nauta, 1972) or a "central autonomic network" (Loewy, 1981). These studies revealed a network of multisynaptic relays descending from the hypothalamus and midbrain to preganglionic neurons in the brainstem and spinal cord. Similar projections, both direct and relayed through the hypothalamus, were also found for a number of limbic system nuclei, including most prominently the amygdala (Nauta, 1972). Neural tracers, immunocytochemical techniques, and electron microscopy subsequently have increased the resolution of the maps of this circuitry and its transmitters (e.g., Loewy, 1981).

The paraventricular nucleus of the hypothalamus is a prototype of the extensive and profound central autonomic coordination. Parvocellular neurons in the paraventricular nucleus project monosynaptically to preganglionic neurons in the dorsal motor nucleus of the vagus and preganglionics in the intermediolateral column of the spinal cord. The paraventricular nucleus also projects to the visceral afferent relay nuclei associated with each of these efferent outflows (the nucleus of the solitary tract and spinal lamina V, respectively). These descending projections influence several cardiovascular and gastrointestinal responses. Illustrating the extent of the central integration of responses, additional neurons of the paraventricular nucleus control corticotropin-releasing factor and oxytocin neuron responses,

BOX 34.4

CENTRAL ANS CIRCUITS INTEGRATE AUTONOMIC REFLEXES WITH AFFECTIVE OR EMOTIONAL RESPONSES

The autonomic nervous system participates in emotion and motivation. Consider Cannon's prototypical fight-or-flight example: The somatic and autonomic adjustments in this emergency situation contain elements we associate with anger or fear. The central relays, including the limbic circuitry just discussed, organize the autonomic components of these responses to external environmental stimuli and coordinate them with the appropriate somatic responses. Thus, the central ANS circuitry is involved in the *expression* of emotional reactions. This central circuitry has been called the "emotional motor system."

Some have hypothesized that in addition to being involved in the expression of emotions, the ANS is also the substrate for the *experience* of emotions. The classic theory of this type was suggested over a century ago independently by James and Lange, who speculated that afferent experience resulting from autonomic adjustments might be the basis of emotional experiences. For the fight-or-flight example, the James–Lange idea could be considered the proposition that "you do not run because you are afraid; you are afraid because you run (and mobilize the requisite autonomic adjustments registered by your visceral afferents)." Cannon, in the early twentieth century, offered an influential refutation of the James–Lange theory and delayed acceptance of these ideas, but most of his argument was based on now-outmoded ideas about the ANS and on a lack of distinction between expression and experience.

In an experiment designed to examine the James–Lange idea, Hohmann (1962) surveyed army veterans who had sustained spinal cord injuries at different levels. Consistent with a prediction of the theory, Hohmann found that, for the veterans' self-reported experiences of both fear and anger, the higher their spinal cord lesions (and thus the more of their visceral afferent inflow that was disconnected from the brain), the greater the reduction in their affective experiences.

In recent years, observations with fMRI and other scanning techniques, with modern electrophysiological techniques combined with tracing and with neuropsychological analyses of patients with brain damage or spinal injury, have converged to suggest that emotions or "feelings" are central representations of visceral afferent inputs. In particular, experiments that elicit visceral afferent activity and others that generate experiences judged to be emotions are associated with comparable activation patterns in limbic structures and forebrain sites. Recently, in views compatible with the James-Lange account of emotion, Damasio (1999), Craig (2002), and others as well have suggested that emotional experience is to the autonomic adjustments of the body what proprioception and somatosensation are to skeletal nervous system operations.

Terry L. Powley

which often are activated or modulated in association with autonomic functions.

Most mesencephalic, diencephalic, and telencephalic nuclei considered part of the limbic system affect visceral motor outflows and are now included in the concept of the central visceromotor system. In addition to the central gray and paramedian regions of the mesencephalon and the entire hypothalamus, including the preoptic hypothalamus, these limbic sites include the amygdala, bed nucleus of the stria terminalis, septal region, hippocampus, cingulate cortex, orbital frontal cortex, and insular and rhinal cortexes.

Summary

Multiple subcortical and diencephalic structures impose special controls over the SNS and PSNS. This network of hierarchically organized central stations, many of them part of the limbic system, forms a central visceral neuroaxis that receives inputs from visceral afferents and issues descending projections to the ANS preganglionic neurons in the brainstem and spinal cord. This central autonomic circuitry organizes and sequences sets of separate autonomic reflexes, coordinates SNS and PSNS responses so that they are synergistic, cross-links autonomic and skeletal responses, and integrates autonomic activity with ongoing and anticipated behavior of the individual.

PERSPECTIVE: FUTURE OF THE AUTONOMIC NERVOUS SYSTEM

Our understanding of the autonomic nervous system is incomplete and evolving, driven by the technological advances in neuroscience. Molecular biology, immunocytochemistry, electron microscopy, fMRI and similar imaging protocols, and other modern techniques are changing the autonomic nervous system definition that was derived with the techniques available to Langley and his contemporaries at the beginning of the twentieth century. The distinction that sympathetic postganglionic neurons are noradrenergic whereas parasympathetic postganglionics are cholinergic has been blurred by the recognition of nonadrenergic noncholinergic (NANC) neurons, nitric oxide synthetase (NOS)-containing neurons, purinergic synapses, a large number of colocalized and coreleased neuropeptides, postganglionic neurons changing their neurotransmitter phenotypes, and other exceptions that broaden the view developed by Langley's pharmacological studies. The canon that the ANS is strictly a motor system without an afferent counterpart is challenged by many who argue on functional and molecular biological grounds that "visceral afferents" belong to the ANS (see Box 34.2).

The proposal that functional differences between the SNS and PSNS divisions of the ANS can be attributed to their contrasting patterns of projection (the sympathetic system preganglionics projecting to postganglionics in a one-to-many pattern; the parasympathetic system projecting in a one-to-few pattern) has not been substantiated by modern analyses of divergence (Wang et al., 1995). Continued application of modern technologies and the prospect of continuing developments promise to define an autonomic nervous system quite different from that envisioned by Langley.

Summary and General Conclusions

The autonomic nervous system is the neural circuitry that maintains homeostasis and health and that enables behavior. It is responsible for the physiological adjustments that achieve what Cannon recognized as the "wisdom of the body." With its sophisticated motor repertoire and complexes of reflexes, including local axon reflexes, segmental or oligosynaptic reflexes, and polysynaptic suprasegmental cascades, the ANS continuously adjusts and defends the body's physiology. The importance of these processes was summarized succinctly by Nauta and Feirtag (1986): "Life depends on the innervation of the viscera; in a way all the rest is biological luxury." The processes go on, for the most part, without awareness or cognitive representations.

The hub of the ANS consists of two separate motor outflows, each hierarchically organized with preganglionic neurons stationed in the CNS and postganglionic neurons located in peripheral ganglia. The sympathetic, or thoracolumbar, division facilitates the mobilization of energy, increases catabolism, and promotes physiological responses that support activity, including emergency responses such as "fight" or "flight." The parasympathetic, or craniosacral, division facilitates conservation of energy, increases anabolism, and supports physiological responses that typically promote rest, digestion, and restoration of body reserves. The adrenergic phenotype of sympathetic postganglionic neurons, the cholinergic phenotype of parasympathetic neurons, and distinguishing complements of neuropeptides and receptor specializations provide neurochemical coding for the two divisions of the ANS. A key peripheral pole—and third distinct division of the ANS—is the enteric nervous system found in the wall of the digestive tract.

The motor outflows of the ANS are efferent limbs of reflexes triggered by visceral and, in some cases, somatic afferents. These reflex circuits have second-order visceral afferent nuclei located adjacent to spinal and medullary nuclei of preganglionic motor neurons (laminae I and V and the nucleus of the solitary tract, respectively).

Higher-order CNS circuitry, including the hypothalamus, limbic system, and a variety of cortical sites, provides

hierarchical control and integration of autonomic reflexes. This hierarchy is responsible for coordinating different autonomic reflexes, integrating autonomic function with somatic activity, and executing response programs that anticipate needs and regulate physiology over more extended time scales than those represented by isolated reflexes.

References

Amenta, F. (Ed.). (1993). *Aging of the autonomic nervous system*. Boca Raton, FL: CRC Press.

Bannister, R. (1989). *Autonomic failure: A textbook of clinical disorders of the autonomic nervous system* (2nd ed.). Oxford: Oxford Medical Publications.

Bieger, D., & Hopkins, D. A. (1987). Viscerotopic representation of the upper alimentary tract in the medulla oblongata in the rat: Nucleus ambiguus. *Journal of Comparative Neurology, 262,* 546–562.

Brading, A. (1999). *The autonomic nervous system and its effectors*. Oxford: Blackwell Science.

Cannon, W. B. (1939). *The wisdom of the body* (2nd ed.). New York: Norton.

Cervero, F., & Foreman, R. D. (1990). Sensory innervation of the viscera. In A. D. Loewy & K. M. Spyer (Eds.), *Central regulation of autonomic functions* (pp. 104–125). New York: Oxford University Press.

Costa, M., Furness, J. B., & Llewellyn-Smith, I. J. (1987). Histochemistry of the enteric nervous system. In L. R. Johnson (Ed.), *Physiology of the gastrointestinal tract* (2nd ed., Vol. 1, pp. 1–40). New York: Raven Press.

Craig, A. D. (2002). How do you feel? Introception: The sense of the physiological condition of the body. *Nature Neuroscience Reviews, 3,* 655–666.

Damasio, A. (1999). *The feeling of what happens: body and emotion in the making of consciousness*. New York: Harcourt Brace.

Dembowsky, K., Czachurski, J., & Seller, H. (1985). Morphology of sympathetic preganglionic neurons in the thoracic spinal cord of the cat: An intracellular horseradish peroxidase study. *Journal of Comparative Neurology, 238,* 453–465.

Fox, E. A., & Powley, T. L. (1992). Morphology of identified preganglionic neurons in the dorsal motor nucleus of the vagus. *Journal of Comparative Neurology, 322,* 79–98.

Furness, J. B. (2006). *The enteric nervous system*. Malden, Massachusetts: Blackwell Publishing.

Hohmann, G. W. (1962). Some affects of spinal cord lesions on experienced emotional feelings. *Psychophysiology, 3,* 143–156.

Holst, M. -C., Kelly, J. B., & Powley, T. L. (1997). Vagal preganglionic projections to the enteric nervous system characterized with PHA-L. *Journal of Comparative Neurology, 381,* 81–100.

Jänig, W. (1996). Neurobiology of visceral afferent neurons: Neuroanatomy, functions, organ regulations and sensations. *Biological Psychology, 42,* 29–51.

Landis, S. C., & Keefe, D. (1983). Evidence for transmitter plasticity in vivo: Developmental changes in properties of cholinergic sympathetic neurons. *Developmental Biology, 98,* 349–372.

Langley, J. N. (1921). *The autonomic nervous system part I*. Cambridge: Heffer.

Loewy, A. D. (1981). Descending pathways to sympathetic and parasympathetic preganglionic neurons. *Journal of the Autonomic Nervous System, 3,* 265–275.

Loewy, A. D., & Spyer, K. M. (Eds.), (1990). *Central Regulation of Autonomic Functions*. New York: Oxford University Press.

Lundberg, J. (1996). Pharmacology of cotransmission in the autonomic nervous system: Integrative aspects on amines, neuropeptides, adenosine triphosphate, amino acids and nitric oxide. *Pharmacological Reviews, 48,* 113–178.

Morgan, C., Nadelhaft, I., & de Groat, W. C. (1981). The distribution of visceral primary afferents from the pelvic nerve within Lissaure's tract and the spinal gray matter and its relationship to the sacral parasympathetic nucleus. *Journal of Comparative Neurology, 201,* 415–440.

Nauta, W. J. H. (1972). The central visceromotor system: A general survey. In C. H. Hockman (Ed.), *Limbic system mechanisms and autonomic function* (pp. 21–40). Springfield, IL: Thomas.

Nauta, W. J. H., & Feirtag, M. (1986). *Fundamental neuroanatomy*. New York: Freeman.

Potter, D. D., Landis, S. C., & Furshpan, E. J. (1980). Dual function during development of rat sympathetic neurones in culture. *The Journal of Experimental Biology, 89,* 57–71.

Strack, A. M., Sawyer, W. B., Marubio, L. M., & Loewy, A. D. (1988). Spinal origin of sympathetic preganglionic neurons in the rat. *Brain Research, 455,* 187–191.

Thomas, L. (1974). *Autonomy: The lives of a cell* (pp. 64–68). New York: Viking Press.

Wang, F. B., Holst, M. -C., & Powley, T. L. (1995). The ratio of pre- to postganglionic neurons and related issues in the autonomic nervous system. *Brain Research Reviews, 21,* 93–115.

Wood, J. D. (1987). Physiology of the enteric nervous system. In L. R. Johnson, J. Christensen, M. J. Jackson, E. D. Jacobson & J. H. Walsh (Eds.), *Physiology of the gastrointestinal tract* (2nd ed., Vol. 1, pp. 67–110). New York: Raven Press.

Suggested Readings

Björklund, A., Hökfelt, T., & Owman, C. (Eds.), (1988). *Handbook of chemical neuroanatomy* (Vol. 6). Amsterdam: Elsevier Science.

Brading, A. (1999). *The autonomic nervous system and its effectors*. Oxford: Blackwell Science.

Burnstock, G. (1992–2000). *The Autonomic nervous system* (Vols. 1–12). Chur, Switzerland: Harwood Academic.

Cameron, O. G. (2002). *Visceral sensory neuroscience interoception*. New York: Oxford University Press.

Cannon, W. B. (1939). *The Wisdom of the body* (2nd ed). New York: Norton.

Gabella, G. (1976). *Structure of the autonomic nervous system*. London: Chapman & Hall.

Hockman, C. H. (Ed.), (1972). *Limbic system mechanisms and autonomic function*. Springfield, IL: Thomas.

Jänig, W. (2006). *The Integrative action of the autonomic nervous system: Neurobiology of homeostasis*. Cambridge: Cambridge University Press.

Ritter, S., Ritter, R. C., & Barnes, C. D. (Eds.), (1992). *Neuroanatomy and physiology of abdominal vagal afferents*. Boca Raton, FL: CRC Press.

Terry L. Powley

This page intentionally left blank

CHAPTER

35

Neural Control of Respiratory and Cardiovascular Functions

Mammals have developed a highly efficient system for tissue delivery of O_2 and removal of CO_2 to support their high metabolic rate. Getting O_2 to the tissues is a two-stage process: a reciprocal pump driven by respiratory muscles exchanges air from outside the body and the lung and a double circulating pump driven by cardiac muscle drives blood through the lungs, where O_2 enters and CO_2 leaves the pulmonary circulation, and then throughout the body, where O_2 leaves and CO_2 enters the systemic circulation. In order to avoid the brain injury that can be triggered by even brief O_2 deprivation, the respiratory and cardiac pumps must work continuously in mammals, from birth until death, without lapses of more than a few minutes.

The role of the brain in assuring O_2 delivery is preeminent. For respiration, the brain generates the rhythmic motor outflows that drive the contraction and relaxation of skeletal muscles that pump the lung and modulate the tone of skeletal and smooth muscles in the upper airways and bronchi to control resistance to air flow. For the circulation, the brain generates the autonomic motor activities that maintain vascular resistance and arterial pressure and that modulate cardiac contractility and heart rate to ensure an appropriate perfusion of all the tissues under a wide range of whole body and local metabolic demands. Through these mechanisms, the brain supports body function over a twenty-fold range in metabolism (in humans from ~0.25 liter × min^{-1} O_2 consumed at rest to ~3–6 liter × min^{-1} O_2 consumed during maximal exercise). Although respiratory muscles account for ≤5% of body metabolism at rest, the brain must drive them efficiently and without serious respiratory muscle fatigue. During exercise in humans, for instance, the brain must increase air exchange from a resting value of 5 up to 200 liters × min^{-1} (maximal performance in world-class athletes) while minimizing the energy costs of breathing to maximize the O_2 available to other working muscles. This chapter summarizes the distinct neural mechanisms underlying the brain's key roles in controlling respiratory and cardiovascular functions and then describes how these markedly different systems work in tandem.

The drive for continuous ventilation in humans is strong. Typically, breathing is the last consequential movement (the "dying" gasp) to disappear following generalized depression of higher brain function, such as during surgical anesthesia or following insults to the brain—for example, hypoglycemic coma or global brain anoxia following a heart attack. Automatic, homeostatic breathing can remain in people who have lost cortical but maintain brainstem and spinal cord function, resulting in ethical and emotional dilemmas in individuals considered "brain dead."

Neuroscientists take advantage of this robustness of breathing to study synaptic, neural, and network mechanisms in anesthetized or decerebrate mammals that continue to breathe. These mechanisms are so robust and sufficiently self-contained that a thin *in vitro* tissue

© 2013 Elsevier Inc. All rights reserved.

slice (~500 μm thick) at an appropriate medullary level in neonatal or late fetal rodent brainstem continues to generate respiratory-related patterns of motor nerve activity.

The cardiovascular system, in addition to carrying O_2 and CO_2, delivers nutrients and removes waste products from the diverse and widely separated tissues of the body. The circulatory system is a conduit for hormonal and metabolic signals arising in the hypothalamus, adrenal cortex, and gut, and for immune cells from, for example, bone marrow, to sites of infection or injury. The central neural regulation of the cardiovascular system provides the overarching coordination among the various aspects of cardiovascular function required not only to maintain a homeostatic cellular environment for all tissues at rest, but especially to meet the enhanced tissue demands during the challenges of behavior, disease, and environmental adversity. The brain regulates the cardiovascular system primarily through the sympathetic and the parasympathetic divisions of the autonomic nervous system (see Chapter 33) and secondarily via hormones, in particular the sympathetically-driven secretion of epinephrine from the adrenal gland.

BREATHING

A Brief Neuroscience-Oriented History

Although the purpose of breathing—to move air into and out of the lungs—was not established until the late eighteenth century, the sites within the body that control breathing have been a constant source of speculation in Western culture (Feldman, 1986). Perhaps the earliest suggestion that the brain was involved came from Galen (ca. A.D. 131–201), who observed that gladiators and animals injured below the neck continued to breathe, but those injured in the neck stopped breathing. Lorry (1760) showed that cerebellectomized rabbits continued to breathe and suggested that the critical circuits lay within the brainstem and upper spinal cord, essentially the contemporary view. Legallois (1813) "extracted" brain tissue to determine what was necessary for breathing; presciently, he localized critical regions to the rostral ventrolateral medulla, close to the *preBötzinger Complex* (preBötC; Figs. 35.2, 35.7), which we recognize today as a key brain site for generating respiration.

Ramon y Cajal (1909) provided an anatomic rationale for the critical role of the medulla in breathing. Examining the afferent/efferent projections of respiratory-related nerves, he focused on two brainstem nuclei: the nucleus of the solitary tract (NTS), the target of pulmonary afferents, and the nucleus ambiguus, which contains cranial motoneurons that innervate upper airway muscles. His network model for breathing is prescient: afferents containing information about lung volume or blood gases combined with intrinsic properties of brainstem neurons to produce a rhythmic outflow to spinal and cranial respiratory motoneurons (Fig. 35.1). In brainstems isolated from goldfish, Adrian and Buytendijk (1931) recorded slow, rhythmic potentials that were similar in timing to the fish's gill movements when intact, physiological evidence that the brainstem contained the critical circuits. Gesell and colleagues (1936) subsequently recorded individual medullary neurons discharging bursts of activity in phase with the breathing rhythm. This finding initiated a concerted effort to identify neurons that generate respiratory rhythm (Box 35.1).

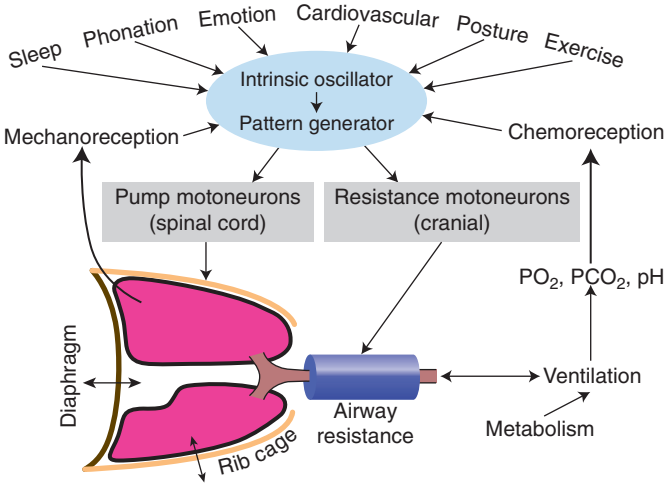


FIGURE 35.1 Functional organization of the CNS control of breathing. Circuitry centered within the medulla oblongata of the brainstem (blue oval) generates an oscillating inspiratory–expiratory rhythm. Neurons within the oscillator circuit generate rhythmic respiratory motor output without requiring sensory signals related to lung inflation. The rhythm is relayed to networks of premotor and interneurons constituting pattern generators that sculpt the detailed firing patterns relayed to spinal and cranial motoneurons. Spinal respiratory motoneurons innervate skeletal muscles, including the diaphragm and muscles of the rib cage, producing the pumping action of the lungs. Cranial motoneurons, e.g., hypoglossal (cranial nerve XII) motoneurons innervating the tongue and vagal motoneurons (cranial nerve X) innervating the glottis, trachea, and bronchi also receive respiratory input that adjusts airway caliber (and hence the resistance to airflow) over the course of each respiratory cycle. The levels of oxygen (PO_2) and carbon dioxide (PCO_2) are sensed by chemoreceptors located in the brain and major arteries. If a low PO_2 or elevated PCO_2 is detected, there is an increase in afferent signals sent to the oscillator and pattern generating circuits to increase breathing. Also, mechanoreceptors located in the airways sense the magnitude of lung inflation and send signals to the central rhythm and pattern generating circuits to modulate the pattern of breathing. Maintenance of appropriate ventilation also requires that breathing be adjusted during a variety of behaviors such as phonation and sleep as well as in response to exercise or changes in posture. These adjustments are mediated via inputs to brainstem rhythm and pattern forming circuits.

BOX 35.1

EXPERIMENTAL ANALYSIS OF VENTILATION

Neuroscientists are constrained by what they can measure. Noninvasive experiments in humans reveal the phenomenology of breathing—for example, the relationship between ventilation and blood CO_2 levels. Brain imaging may reveal which regions are involved, but at present, spatiotemporal resolution is too coarse to show networks, much less neurons or synapses. Respiration lends itself to study because mammals suitable for experimentation breathe much like humans. Breathing continues following anesthesia or decerebration, and the neural patterns remain following paralysis (and mechanical ventilation), making possible experimental procedures requiring highly invasive techniques, such as single neuron recording. Compared to studies of neurons in culture or in tissue slices, experiments in whole mammals have serious limitations: (1) movements of the heart and lungs make certain types of single neuron recording, as well as modern imaging techniques such as fMRI, difficult in the brainstem and spinal cord; (2) the blood–brain barrier is intact, making precise control of brain extracellular fluid impossible; and (3) acute experiments in large mammals, such as cats, dogs, and goats, are expensive and can last 36 h or more, requiring the dedicated and concurrent efforts of several investigators. Fortunately, the robust features of breathing that allow it to persist when other behaviors are suppressed by anesthesia or decerebration permit further reduction in the experimental preparation that removes these limitations. Moreover, a respiratory-related rhythm persists in the brainstem and spinal cord removed from a neonatal rodent and even in a particular slice of brainstem isolated from this preparation. Using such slices, one can correlate exquisite measures of a neuron's intrinsic and synaptic properties and projections with measures of endogenous behavior, such as its firing pattern. With many common and otherwise useful *in vitro* preparations, such as the hippocampal slice, such correlations cannot be made because no endogenous behavior *in vitro* relates to known behavior *in vivo*. In general, en bloc *in vitro*, or slice, preparations play an important role in the study of basic cellular, synaptic, and integrative mechanisms underlying respiratory motor behavior.

Jack L. Feldman and Donald R. McCrimmon

WHERE ARE THE NEURONS GENERATING THE RESPIRATORY PATTERN?

Three distinct functional classes of respiratory neurons underlie the generation of the discharge pattern on the motor nerves innervating skeletal muscles of the respiratory pump and airways: respiratory muscle motoneurons, rhythm-generating neurons, and premotoneurons that link and coordinate rhythm-generating neurons to motoneurons and transform that rhythm into a mechanically efficient pattern of respiratory muscle activity. Nonrhythmic neurons that modulate the behavior of these respiratory neurons are also important. These include neurons involved in sensing blood and brain levels of O_2 and CO_2—that is, chemoreceptors—or may reflect sleep-wake state, such as serotonergic or (nor)adrenergic neurons. Brainstem respiratory neurons are concentrated within three regions (Fig. 35.2).

The *dorsal respiratory group (DRG)* is located within the caudal half of the NTS, the primary target of afferent input from the glossopharyngeal and vagus nerves. DRG neurons are optimally located to receive respiratory-related sensory input. DRG neurons provide input to other groups of respiratory neurons within the *ventral respiratory column* (VRC) of the ventral medulla, the *pontine respiratory group* (PRG), as well as respiratory motoneurons in the spinal cord.

The VRC forms a continuous rostrocaudal field in the ventrolateral medulla with multiple, physiologically distinct subregions, extending from the facial nucleus to the spinomedullary junction. The caudal half of the VRC, termed the ventral respiratory group (VRG), contains respiratory-modulated bulbospinal premotoneurons. Inspiratory neurons in the rostral VRG (rVRG) provide mainly excitatory premotor drive to interneurons projecting to inspiratory muscle motoneurons, including phrenic motoneurons innervating the diaphragm (the principal inspiratory muscle) in the midcervical (C3-C5 in humans) spinal cord and external intercostal muscles of the rib cage in the thoracic spinal cord. Expiratory bulbospinal neurons in the caudal VRG (cVRG) are mainly excitatory and project to interneurons innervating spinal motoneurons that drive internal intercostal and abdominal expiratory pump muscles in the thoracic and lumbar spinal cords, respectively. Inhibitory expiratory premotor neurons in the cVRG and in the *Bötzinger complex* (BötC) inhibit spinal inspiratory motor pools.

In the rostral VRC, rostral to the rVRG, are the preBötC (an essential structure for normal rhythm generation); the BötC; the *retrotrapezoid nucleus* (RTN) that contains central chemosensory cells whose activity increases in parallel with increasing local tissue acidity and provides an important source of respiratory drive; and finally, the *parafacial respiratory group* (pFRG) which plays an important role

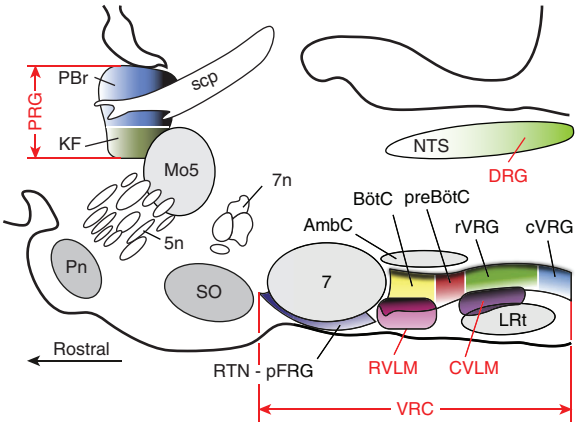


FIGURE 35.2 Respiratory and cardiovascular cell groups in a sagittal view of the rat brainstem. Functionally distinct brainstem compartments related to breathing or cardiovascular control are indicated by shaded regions of different colors. Red text indicates the three primary respiratory and two primary cardiovascular cell groups. Subregions of two of the primary respiratory groups, the ventral respiratory column (VRC) and pontine respiratory group (PRG), are indicated with black labeling. Nearby brainstem structural landmarks are indicated in gray for prominent neuronal nuclei, or outlined without shading for major fiber bundles. 7, facial nucleus; 7n, facial nerve; AmbC, compact part of nucleus ambiguus; BötC, Bötzinger complex; cVRG, caudal division of ventral respiratory group; DRG, dorsal respiratory group; KF, Kölliker–Fuse nucleus; LRt, lateral reticular nucleus; Mo5, motor nucleus of the trigeminal nerve; NTS, nucleus of the solitary tract; PBr, parabrachial region; Pn, basilar pontine nuclei; preBötC, preBötzinger complex; RTN - pFRG, retrotrapezoid nucleus and parafacial respiratory group; rVRG, rostral division of ventral respiratory group; scp, superior cerebellar peduncle; SO, superior olive; VRG, ventral respiratory group. *Modified from Alheid and McCrimmon (2008).*

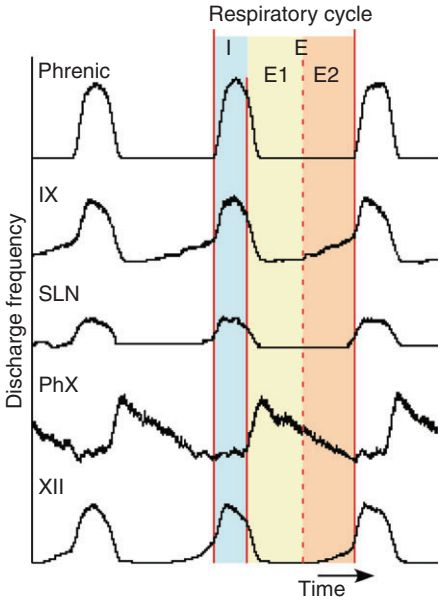


FIGURE 35.3 A diverse pattern of respiratory motor activity is exhibited on spinal (phrenic) and cranial nerves innervating the diaphragm and airways in an anesthetized rat. Low-pass filtered traces of nerve activity show the overall pattern of discharge. Red vertical lines indicate phase transitions between inspiration (I) and expiration (E), which itself is divided into two subphases: E1 (postinspiration) and E2. Note that the onset of inspiratory activity on cranial nerves precedes the onset of activity on the phrenic nerve. IX, glossopharyngeal nerve; PhX, pharyngeal branch of vagus nerve; SLN, superior laryngeal nerve XII, hypoglossal nerve. Traces depict activity in a rat, but activity is similar in other mammals although respiratory frequency varies from greater than 100 × min^{-1} in mice to about 3 × min^{-1} in giraffes. *Adapted from Hayashi and McCrimmon (1996).*

in the generation of expiratory motor output (Feldman & Del Negro, 2006).

The PRG in the rostral dorsolateral pons encompasses portions of the parabrachial complex and the ventrally adjacent Kölliker-Fuse nucleus. PRG neurons may be necessary for the full expression of the normal breathing pattern, but do not seem essential for rhythm generation *per se*. Further, they contribute to the reflex changes in breathing elicited by sensory afferents sensitive to lung volume. They also contribute to the reflex changes in breathing induced by input from chemoreceptors sensitive to the levels of arterial blood O_2, CO_2, and pH. PRG neurons may also contribute to the coordination of respiration with the cardiovascular system, and mediate the effects of breathing rhythm on suprapontine structures and vice versa.

DISCHARGE PATTERNS OF RESPIRATORY NEURONS

Although inspiration and expiration constitute two obvious phases of the mammalian respiratory cycle, the discharge patterns of inspiratory (diaphragm) and expiratory—for example, internal intercostal—muscle activity suggest two expiratory phases. During the first expiratory phase (E1, termed postinspiration), the diaphragm often exhibits a low level of "postinspiratory" activity that slows diaphragm lengthening at the onset of expiration, thus reducing the initial rate of expiratory airflow (Fig. 35.3). Postinspiratory activity typically declines rapidly, ending by midexpiration. In a healthy individual under resting conditions there is little or no activity of expiratory pump muscles, with expiration occurring via the passive recoil of the lung. At higher levels of ventilation, such as during exercise (note: here exercise means sustained activity of large muscles, such as would be involved in fleeing a predator or chasing prey, or running on a treadmill!), an augmenting pattern of expiratory muscle activity in the second (E2) phase increases the expiratory airflow. These motor patterns are generated by interactions among respiratory neurons primarily localized in the respiratory cell groups illustrated in Figure 35.2.

MODELS FOR RESPIRATORY PATTERN GENERATION

Oscillations such as the respiratory motor patterns shown in Figure 35.3 are a common form of dynamic activity in physical and biological systems that have

attracted modelers since the early seventeenth century when philosopher René Descartes used analytic reasoning to model the mind. Since the 1960s, computational models to explain how rhythm-generating neurons in the preBötC, RTN/pFRG interact with the rest of the network, including premotor and motoneurons, to produce appropriate motor patterns have successfully reproduced human and animal breathing patterns. These models are useful for summarizing interaction mechanisms among components of the respiratory system, but modeling efforts still seek the elusive goal of nontrivial, testable predictions. Although current modeling elucidates plausible mechanisms, further data on network connections, intrinsic membrane properties, and synaptic interactions are needed so models are sufficiently constrained to provide novel insights into the underlying mechanisms.

A Respiratory Rhythm Generator is Discovered in the preBötC

Since the late eighteenth century, the brainstem has been recognized as the source of respiratory rhythm, but only in the past 25 years has the precise locale of the obligate structures been determined. A critical step in this advance was the development of novel *in vitro* preparations (see Box 35.1). In the 1980s, Suzue and colleagues (Suzue, 1984) noted that the isolated neonatal rat brainstem and spinal cord, termed the "en bloc" preparation, generated a respiratory rhythmic motor output in spinal nerves that innervate muscles of the respiratory pump, as well as in cranial nerves innervating the airways, which control airway resistance. This *in vitro* preparation permitted a refinement of the "blunt spatula" brainstem transection approach used almost 200 years earlier by Legallois (1813) to identify structures critical for respiratory rhythm generation. In the contemporary case, removing thin (50–75 μm) sections from either rostral or caudal ends of the *en bloc* preparation, while recording respiratory activity from spinal or cranial nerves resulted in the isolation of a small, ventral medullary region, named the preBötC, as an essential component in the rhythm generating circuitry (Smith, Ellenberger, Ballanyi, Richter, & Feldman, 1991). Several experimental observations indicate that neurons within the preBötC provide a minimum essential circuitry for respiratory rhythm generation (Feldman & Del Negro, 2006). (1) Slices containing the preBötC generate a rhythm indistinguishable from the respiratory-related rhythm *en bloc*, although this pattern differs from that seen *in vivo*. (2) In intact adult *awake* rats, neurotoxic lesions of a subset of preBötC neurons containing the receptor for substance P (i.e., neurokinin-1 receptor), irreversibly induces a pathological, ataxic breathing pattern quite different from normal breathing (Fig. 35.4 and see Box 35.2). (3) Essentially normal inspiratory motor pattern persists after transection of the rat brainstem just rostral to the

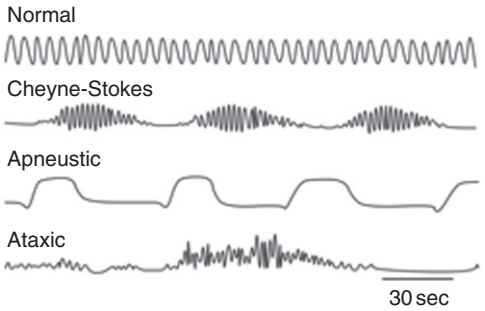


FIGURE 35.4 Abnormal breathing patterns resulting from CNS disorders. The ordinate is lung volume. *Adapted from Plum and Posner (1980).*

BOX 35.2

CNS LESIONS PRODUCE ABNORMAL BREATHING PATTERNS

Many brain injuries and diseases produce abnormal breathing patterns (Plum & Posner, 1980). Because many brain regions provide afferent input to the neurons generating the breathing rhythm, pathology in regions not normally associated with the generation of breathing can produce abnormal breathing patterns. Despite the diffuse nature of many pathologies that give rise to abnormal breathing patterns, Plum and colleagues systematically characterized several breathing disorders (Fig. 35.4) arising from specific CNS pathologies.

Apneustic breathing is marked by prolonged inspiratory periods. In humans, the most frequent pattern is inspirations lasting 2–3 s alternating with prolonged expiratory pauses. In cats, the prolonged inspiratory periods are associated with plateaus in inspiratory drive that can last minutes (leading to death in the absence of mechanical ventilation). Apneusis is observed in people with lesions of the pons, including, or just ventral to, the pontine respiratory group. In experimental animals, apneusis requires not only lesions of the pontine respiratory group, but also interruption of vagal afferent input.

Lesions of the corticobulbar or corticospinal tracts can lead to *Cheyne–Stokes respiration*, a rhythmic waxing and waning of breathing. Periods of no breathing (apnea) follow each period of waning inspiratory depth. Lesions of the corticobulbar and corticospinal tracts can also result in loss of voluntary control of breathing. In pseudobulbar palsy, for example, voluntary control of breathing and of cranial

BOX 35.2 *(cont'd)*

motoneuron function is lost secondary to a lesion often located dorsomedially in the base of the pons.

Extensive bilateral damage to the medullary respiratory groups can severely disrupt or abolish respiratory rhythm, resulting in death unless artificial ventilation is initiated immediately. Fortunately, unilateral damage does not appear to be sufficient to cause severe disruption of respiratory rhythm- and pattern-generating mechanisms. Because two vertebral arteries supply blood to the medulla, bilateral damage from an infarct or embolism is unlikely.

Less extensive damage to medullary respiratory structures can produce *ataxic breathing*, an irregular pattern of breathing with randomly occurring large and small breaths, periods of apnea and low breathing frequency.

Spinal injuries that affect the long axons of bulbospinal premotoneurons that transmit respiratory drive to spinal motoneurons can be life-threatening. Damage between C3 and S1 can diminish the ability to generate inspiratory or expiratory movements of abdominal and intercostal muscles, with the effects more pronounced the more rostral the injury. Injuries between the lower brainstem and C3 will also affect the diaphragm. Such injuries require immediate artificial support of ventilation to maintain life. This type of cervical injury is more common in young, active individuals, resulting from dives into shallow water or being thrown from a horse.

Jack L. Feldman and Donald R. McCrimmon

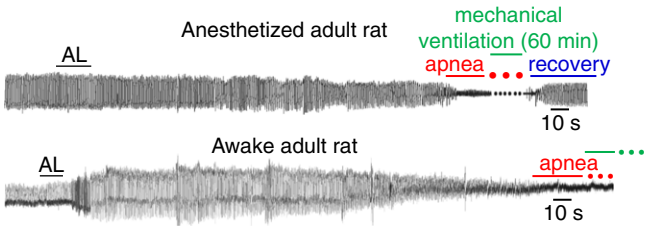


FIGURE 35.5 PreBötzinger complex (preBötC) is essential for breathing in adult rats. Rapid silencing of allatostatin receptor (AlstR)-expressing, preBötC somatostatin (Sst) neurons induces persistent (>45 min when mechanically ventilated) apnea in anesthetized or awake adult rats. Traces are plethysmographic recordings of breathing movements. Allatostatin (AL) administered intracerebrocisternally induces a gradual decline in frequency and tidal volume until apnea develops after several minutes. After ~60 minutes of mechanical ventilation, rats resume spontaneous breathing. The ordinate is lung volume. *From Tan et al. (2008).*

preBötC, which removes all pontine circuits and the RTN/pFRG although phasic expiratory motor activity is abolished. (4) Rapid transient silencing of a subpopulation of preBötC neurons, those glutamatergic neurons that also produce the peptide somatostatin (~500 neurons/side), leads rapidly to a persistent apnea in awake rats (Fig. 35.5). Taken together, these experimental results underlie the now widely accepted view that the preBötC is an essential component of the circuitry generating breathing rhythm.

A Second Respiratory Rhythm Generator is Unveiled

The RTN/pFRG (Fig. 35.2), a region rostral to the preBötC, is now thought to play a critical role in generating respiratory rhythm as well as in its ontogeny and in chemoreception (see Box 35.3). Functional evidence for the presence of this second oscillator initially came from experiments exploiting differences in the opioid sensitivity of preBötC neurons, which are inhibited by opioid agonists, and RTN/pFRG neurons, that appear unaffected. In *in vitro* slices that contain only the preBötC, opioids slow the inspiratory rhythm in a graded fashion, gradually increasing the interval between inspiratory bursts (Fig. 35.6A, top). In the more intact *en bloc* preparation that contains both the preBötC and the RTN/pFRG, opioid agonists also cause a slowing of inspiratory rhythm, but instead of progressively lengthening the interval between inspiratory bursts, some inspiratory bursts are skipped. This results in a stepwise lengthening of the interval between inspiratory bursts (termed "quantal slowing" (Mellen, Janczewski, Bocchiaro, & Feldman, 2003)), with the amount of slowing dependent on how many inspiratory bursts were skipped (Fig. 35.6A, bottom). While this is happening in the motor output, RTN/pFRG expiratory neurons continue to burst at the control frequency with no skips and remain phase-locked to motor nerve output when it occurs, whereas preBötC neurons only burst when there is a motor output (Fig. 35.6B). Quantal slowing of the inspiratory rhythm also occurs *in vivo*, where, the frequency of expiratory motor nerve bursts is unaffected! This is unlikely to be due to a simple block of inspiratory motor output, as the pattern of expiratory motor activity is quite different in cycles with and without inspiration (Fig. 35.6C, right). Excitation or disinhibition of RTN/pFRG neurons can transform a breathing pattern with no active expiratory activity into one with considerable expiratory motor activity (Fig. 35.6D) (Pagliardini, Janczewski, Tan, Dickson, Deisseroth, & Feldman, 2011). Finally, lesions or transections can be used to separate

BOX 35.3

WHY TWO DISTINCT OSCILLATORS?

Two key events characterize the evolution of breathing from fish to mammals. The first includes transformations that support the transition from aquatic to terrestrial breathing in vertebrates. "Breathing" in fish is driven by cranial nerve innervation of the gills. As primitive fish (such as the ancestors of the lungfish) ventured onto land, rhythmic activity of muscles used for other purposes evolved to drive the emerging lung, and this activity might have had a neural origin that was independent of the rhythm driving the gill muscles. This speculative hypothesis is supported by observations in amphibians. Developing frogs have two distinct breathing patterns driven by different muscles and nerves: a buccal pattern that uses cranial nerve-innervated muscles to essentially swallow air and a lung pattern that uses respiratory muscles to pump the lung. Either the buccal or the lung pattern can be suppressed without abolishing its counterpart, but when both are active they appear to be phase-locked. Notably, the lung rhythm is depressed by opioids (like the mammalian preBötC neurons and inspiration), whereas the buccal rhythm is not (like pre-I neurons and expiration). A prediction from this hypothesis is that the preBötC is absent in fish.

The second key event was the emergence of the diaphragm in mammals. Respiratory physiology in vertebrates has evolved to support higher resting and peak ventilation. For example, lizards have gular pumps that uncouple respiration from locomotor musculature, which enhances O_2 consumption to support high-speed locomotion. Birds utilize sacs to pump O_2-rich air through noncompliant lungs, in support of flight. Uniquely, mammals have muscular diaphragms, a profoundly important ventilatory advance as it allows for continuous high basal rates of ventilation and for the very high levels of ventilation necessary to support substantial increases in metabolism. The improvement in gas exchange associated with the development of the diaphragm was probably crucial for brain evolution, as it was a platform for evolution of a brain that consumes extraordinary amounts of O_2 and is generally intolerant of even transient anoxia. The neural circuits driving the primitive diaphragm could have evolved independently of preexisting circuitry driving the rhythmic pumping machinery of the rib cage, abdominal muscles, or upper airway muscles, or it could have been based on the opioid-sensitive oscillator driving lung breathing in amphibia. Nonetheless, a respiratory movement pattern consisting of alternating inspiratory activity to lower the diaphragm with expiratory motor activity has considerable mechanical advantages for efficient ventilation over a broad range of metabolic demand. Therefore, the neural oscillators for breathing, which may have arisen independently, might be expected to evolve to be coupled.

There may be a special role for two oscillators at birth. In mammals there is a surge of endogenous opioids at birth. This postnatal increase in opioid release would depress the excitability of the preBötC but not the RTN/pFRG, thereby allowing the latter to promote breathing immediately after birth, either by rhythmically driving breathing or by providing a drive necessary to keep the preBötC rhythmic. This view is supported by experiments in Krox-$20^{-/-}$ mice that have neuroanatomical defects that result in deletion of the brainstem region containing the RTN/pFRG but do not affect the preBötC. These mutant mice die of fatal apneas within 18 hours after birth, but brief postnatal administration of naloxone (an opioid receptor agonist which has no effect in control mice) rescues these pups by eliminating their apneas. Thus, without the RTN/pFRG to provide respiratory drive, the opioid surge at birth could fatally depress preBötC function. We suggest that in Krox-$20^{-/-}$ mice, in the absence of the RTN/pFRG, the opioid-induced depression of the preBötC causes prolonged and ultimately fatal apneas; naloxone would reverse this depression, which is developmentally relieved within two days, allowing the preBötC to function adequately for survival. This special role for the RTN/pFRG at birth is consistent with the two-oscillator model.

Jack L. Feldman and Donald R. McCrimmon

the regions that generate inspiratory and expiratory activity. In vagotomized and anesthetized juvenile rats, complete brainstem transections rostral to the VII nucleus that leave the RTN/pFRG and preBötC intact do not markedly affect inspiratory or expiratory motor bursting activity, whereas transections that disconnect the rostral portion of the RTN/pFRG from the more caudal preBötC abolish expiratory motor activity without affecting inspiratory rhythm. Additionally, continuous lung inflation suppresses inspiratory activity and enhances expiratory drive, whereas lung deflations increase the rate of inspiration, while minimizing rhythmic expiratory efforts.

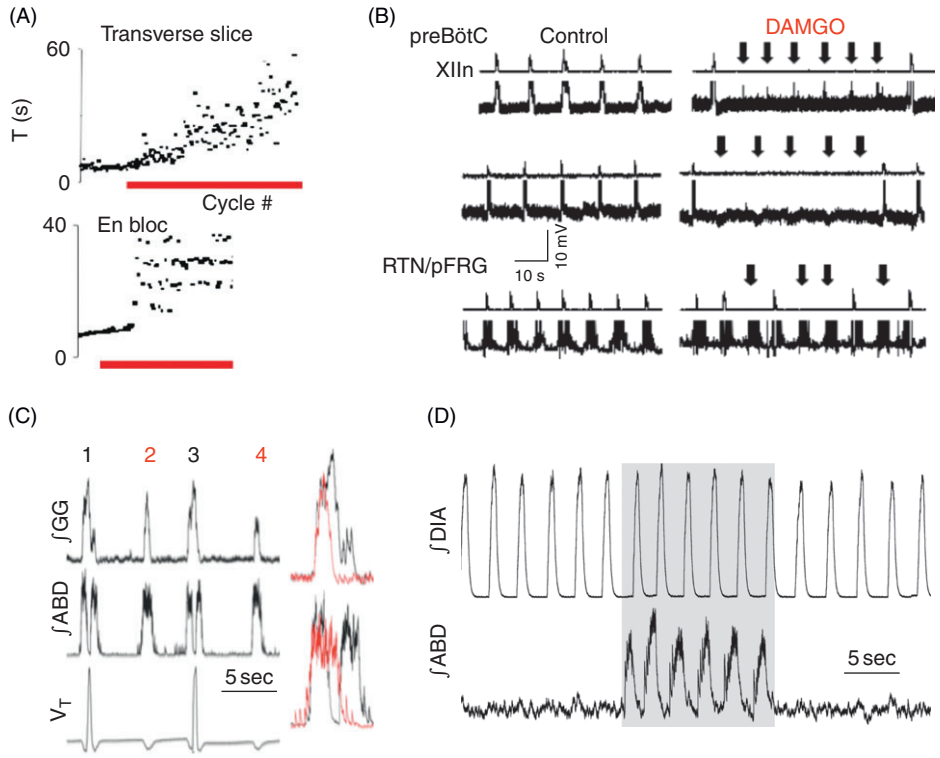


FIGURE 35.6 Evidence for two respiratory oscillators. (A) Slowing of the inspiratory period *in vitro* in a slice (top) and en bloc (bottom), before and after DAMGO, an opiate agonist (red bar): top—continuous slowing of rhythm in the slice; bottom—quantal slowing of rhythm *en bloc*. T(s), time in seconds between successive bursts of inspiratory neuronal activity recorded either from the preBötC (top) or a cervical ventral root (bottom). (B) Simultaneous recordings of inspiratory burst activity in XII nerve and preBötC inspiratory neurons (top two pairs of traces) and pre-I neuron in RTN/pFRG (bottom pair). Arrows in DAMGO traces indicate subthreshold events in preBötC neurons during skipped bursts, which occur at the approximate expected time of inspiratory bursts at control frequency, or unperturbed bursting in pre-I neurons. (C) *In vivo* recordings (juvenile rat) following fentanyl, an opiate agonist, injection. Left: typical recording with normal cycles (1, 3) interspersed with cycles without inspiratory activity (2, 4). Traces from top: integrated genioglossus muscle EMG; integrated abdominal muscle EMG, tidal volume (V_T). Right: Superposition of $\int EMG_{GG}$ and $\int EMG_{ABD}$ in a normal cycle (black) compared to an inspiratory skipped cycle (red). Note the marked differences, especially in the $\int EMG_{ABD}$ burst. (D) Optogenetic stimulation of RTN neurons turns on abdominal (expiratory) nerve activity, with minimal effect on inspiratory activity. *A, B adapted from Mellen et al. (2003); C adapted from Janczewski and Feldman (2006); D adapted from Pagliardini, et al. (2011).*

Summary

The working paradigm for the core engine of breathing is that the preBötC generates inspiratory rhythm and the RTN/pFRG generates expiratory activity, and these two oscillators are coupled (Fig. 35.7).

MECHANORECEPTORS IN THE LUNGS ADJUST BREATHING PATTERN AND INITIATE PROTECTIVE REFLEXES

Many different patterns of muscle activity can produce appropriate ventilation; however, the specific pattern should minimize energy expenditure. Mechanical efficiency, in turn, depends on factors such as posture, and lung and chest wall mechanics; for example, lung injury can lead to an increase in fibrous tissue in the lung, making it harder to inflate than a normal lung, and thus for this condition, rapid, shallow breathing is more efficient. Feedback about the mechanical status of the lungs and chest wall is provided by mechanoreceptors. Protective respiratory reflexes, such as sneezing and coughing, are also generated by pulmonary receptors.

There are three distinct groups of lung afferents, and all of them have axons in the vagus nerve that terminate in the NTS (Kubin, Alheid, Zuperku, & McCrimmon, 2006). Two groups have large, myelinated A fibers. These include slowly adapting and rapidly adapting pulmonary stretch receptors. The third group has unmyelinated C fibers and responds to inhaled irritants such as smoke. Most natural stimuli—for example, pulmonary edema or alveolar collapse—probably activate multiple receptor types.

Slowly Adapting Pulmonary Stretch Receptors

Located in airway smooth muscle, slowly adapting pulmonary stretch receptors (SARs) are activated when the airways stretch during lung inflation (Kubin et al., 2006) providing the central pattern generator with information

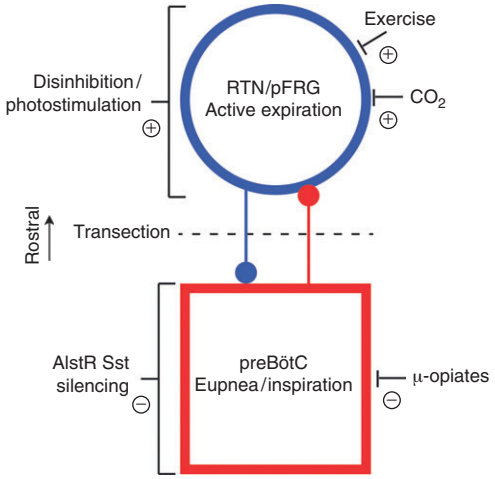


FIGURE 35.7 At the core of current models for respiratory rhythmogenesis are two medullary nuclei, the preBötC and the RTN/pFRG. The preBötC is essential for the generation of inspiratory motor activity, sufficient for breathing at rest. Endogenous stimulation of the RTN/pFRG by elevated CO_2, or perhaps exercise, is hypothesized to produce active expiration (absent at rest in adult mammals). The above figure summarizes results discussed in the text.

about lung volume. Their activation drives reflexes by which lung inflation shortens inspiration and prolongs expiration (Breuer–Hering reflex), which is an important determinant of inspiratory duration during normal relaxed breathing in most mammalian species, with the notable exception of humans. The activation of these receptors shortens inspiration and increases breathing frequency. In humans, activation of SARs by normal breaths only modestly alters inspiratory timing. Their activation also relaxes airway smooth muscle, dilating the airways to make breathing easier.

Rapidly Adapting Pulmonary Stretch Receptors

Located in airways, rapidly adapting pulmonary stretch receptors (RARs) initiate protective reflexes in response to a variety of stimuli, including large or rapid lung inflation or deflation, inhaled irritants, and, possibly, airway edema (Canning & Spina, 2009). Cigarette smoke activates RARs and elicits a cough to rid the airway of the offending material. RARs elicit a larger than normal inspiration (sigh) in response to alveolar collapse (atelectasis), inflating the lung and popping open the collapsed alveoli. Also contributing to cough is a recently identified group of airway receptors, termed cough receptors, with small myelinated fibers (A-∂ fibers).

Bronchopulmonary C Fibers

C fiber afferent activation elicits apnea followed by rapid shallow breathing. Like RARs, C fibers are polymodal, activated by chemical and mechanical stimuli. Their activation also enhances airway mucus secretion. Inhaled particles are trapped in the mucus and removed from the airways by the action of cilia, which continuously (in the nonsmoker) move the mucus and trapped particles toward the mouth to be swallowed or expectorated.

Summary

Mechanoreceptors, especially those in the lungs, are essential for the precise regulation of the timing and amplitude of breathing. They also participate in reflexes, such as those eliciting coughing and mucous secretion, that protect the airways and lungs from compromises in airflow and inhaled irritants.

CARDIOVASCULAR

Organization and Functions of the Innervation of the Cardiovascular System

The heart, unlike skeletal muscle, can beat rhythmically in the absence of its neural input. However, neural inputs are essential for modulating cardiac contractility—that is, force of contraction—and heart rate. Sympathetic and parasympathetic nerves innervate the heart (Fig. 35.8), regulating cardiac function in a push-pull manner. Increasing the sympathetic input to the heart increases heart rate and its contractility, thereby increasing the amount of blood pumped—that is, cardiac output. The sympathetic innervation is via the cardiac nerve that arises from noradrenergic sympathetic ganglion cells (primarily in the stellate ganglion) driven by cholinergic preganglionic neurons in the intermediolateral nucleus of the upper thoracic spinal cord. Increasing the parasympathetic input to the heart decreases the heart rate. The parasympathetic, cholinergic postganglionic neurons are within adipose tissue apposing the heart, driven by cholinergic vagal preganglionic neurons whose somata are in the nucleus ambiguus and the dorsal motor nucleus of the vagus.

Blood vessels, which have only a low contractile tone in the absence of neural input, generally receive only a sympathetic innervation that mediates vasoconstriction (although several notable exceptions are described in the next paragraph). Increased vascular sympathetic activity increases the tone of vascular smooth muscle, resulting in vasoconstriction that increases arterial blood pressure. Decreased vascular sympathetic activity allows the blood vessels to dilate, decreasing blood pressure. Sympathetic input to veins increases venous wall tension, which increases cardiac filling and cardiac output. The sympathetic vascular innervation is from noradrenergic sympathetic ganglion cells in prevertebral (visceral blood

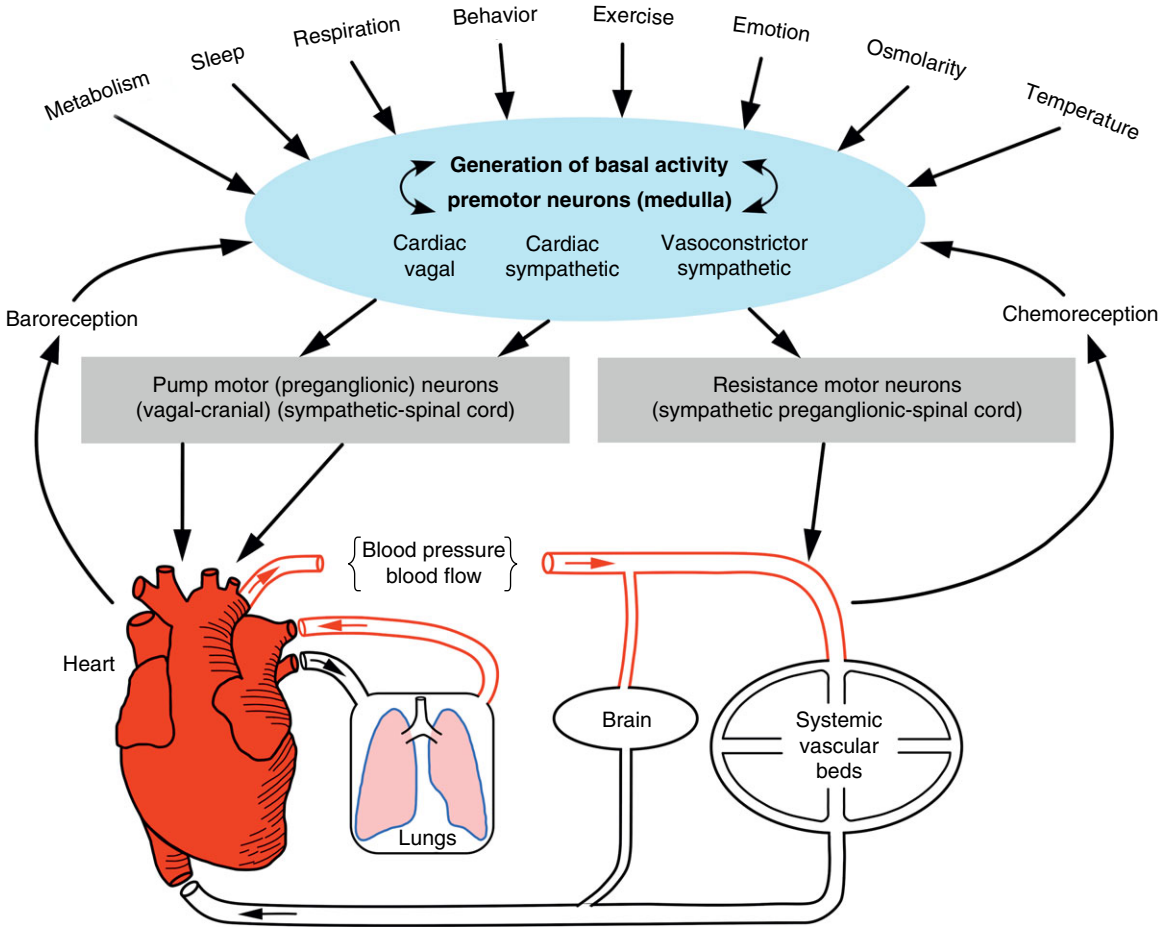


FIGURE 35.8 Functional organization of the CNS regulation of cardiovascular function. The medulla contains sympathetic neurons, including sympathetic premotor neurons that are responsible for the significant level of basal sympathetic nerve discharge to the heart and vasculature. Although the medullary mechanisms underlying the generation of basal sympathetic tone remain to be identified, they result in the activity of medullary sympathetic premotor neurons that provides the essential excitation to spinal sympathetic preganglionic neurons that maintains the basal sympathetic tone to the heart and vasculature. The basal activity of sympathetic nerves maintains a release of norepinephrine in the heart and vascular beds that supports, respectively, the rate and strength of cardiac pumping and the constriction of blood vessels that produces the resistance to blood flow required to maintain blood pressure at a level sufficient to ensure adequate blood flow to the brain, even against the pull of gravity. Selective activation or inhibition of the sympathetic outflow to the vascular beds in different tissues allows relative reductions or increases in blood flow to those tissues as needed to support a wide range of behaviors and of homeostatic adjustments to environmental challenges. Many of the stimuli for changes in sympathetic and vagal outflow to the cardiovascular system, from homeostatic reflexes, to diseases, dehydration, drugs, and metabolism, elicit their responses through inputs, including those from the hypothalamus, to the medullary networks, including the premotor neurons, that generate sympathetic tone.

vessels) and paravertebral (vessels in the skin and skeletal muscle) ganglia close to their target tissues. Their preganglionic neurons are in the intermediolateral nucleus throughout the thoracolumbar spinal cord segments.

Sympathetic activation can be highly localized or global in effect. Sympathetic preganglionic neurons and ganglion cells are target-specific, providing a substrate for selective CNS regulation of the heart and of the blood vessels in each tissue. Chromaffin cells in the adrenal medulla, driven by adrenal sympathetic preganglionic neurons in the thoracic spinal cord, secrete the catecholamines, epinephrine and norepinephrine, into the circulation as part of the stress response. Although normally at a low level, circulating epinephrine stimulates β1 adrenergic receptors on cardiac tissue to increase cardiac output and β2 adrenergic receptors to dilate skeletal muscle blood vessels, both supporting the "fight-or-flight" response to danger, as well as the muscle hyperemia needed for exercise.

There are several notable exceptions to the concept that blood vessels only receive a sympathetic postganglionic innervation. Innervation of arteries supplying the penis or the clitoris includes a parasympathetic component underlying an active vasodilation that maintains penile erection or clitoral engorgement and determines the success of reproductive behavior. Indeed, the role of nitric oxide as the mediator of this parasympathetic vasodilation

is the basis for a variety of successful pharmacological agents to combat erectile dysfunction. Parasympathetic inputs to the eye produce a nitric oxide-mediated vasodilation in the uveal vasculature that contributes both to the fine control of the availability of metabolic fuels to parts of the eye that lack an intrinsic blood supply and to optimizing the retinal image through variations in uveal light absorption. In humans, sympathetic vasodilator nerves to the skin are principally responsible for an acetylcholine-mediated relaxation of vascular smooth muscle that increases cutaneous blood flow for increased heat loss in hyperthermic environments. The parasympathetic innervation of the salivary glands elicits a local vasodilation that supports an increase in salivary secretion. Norepinephrine released from sympathetic nerve terminals innervating brown adipose tissue acts not only on brown adipocytes to induce thermogenesis, but also on local vascular endothelial ß-adrenoceptors to induce nitric oxide synthase and thereby provide the increase in tissue blood flow that distributes the heat throughout the body. Similarly, the increase in adrenal medullary blood flow that accompanies the sympathetic stimulation of adrenal catecholamine release is dependent on a sympathetic preganglionic, cholinergic stimulation of nitric oxide production.

Basal Neural Tone to the Cardiovascular System

The sympathetic and parasympathetic nerves to the heart and the sympathetic nerves to the vasculature all have a basal level of discharge ultimately controlled by central neural circuits innervating the various populations of preganglionic neurons (Fig. 35.8). Due to the intrinsic rhythmicity of cardiac pacemaker cells, the heart beats without extrinsic innervation (as is the case for human heart transplants). Heart rate is modulated by the combined sympathetic, parasympathetic, and adrenal epinephrine effects on cardiac pacemakers. For instance, resting heart rate is slower in trained athletes due to an increased vagal (parasympathetic) tone—a reflex that maintains a constant resting cardiac output in the face of the increased stroke volume resulting from the stronger cardiac muscle produced by training. In contrast, although there are a variety of chemical signals that affect vascular smooth muscle contraction, under normal conditions, the level of sympathetic nerve activity to blood vessels is the primary factor determining vascular diameter and, in turn, the resistance to blood flow throughout the vascular network. Vascular smooth muscle, in the absence of sympathetic input, has insufficient tone to provide enough resistance to keep arterial blood pressure at a level sufficient to provide adequate blood flow to meet the metabolic needs of the brain and other tissues in the body. Thus, a basal level of sympathetic tone is essential.

Sympathetic Premotor Neurons and Generation of Basal Sympathetic Tone

The cardiovascular sympathetic preganglionic neurons receive their primary excitatory drive from premotor neurons in the rostral ventrolateral medulla (RVLM). These bulbospinal neurons (Fig. 35.9) are glutamatergic, but many also contain the enzymes for catecholamine synthesis (C1 cell group). Silencing these neurons with local injections of muscimol (a $GABA_A$ receptor agonist), or transection of their bulbospinal axons as in spinal cord injury, results in a precipitous fall in arterial pressure due to marked reductions in cardiac output and in vascular resistance. RVLM neurons have a significant basal discharge rate that supports basal cardiovascular sympathetic nerve activity. Modulation of this activity underlies a variety of reflex- and centrally-evoked changes in sympathetic nerve activity, including the baroreceptor and chemoreceptor reflexes and the response to physical exercise, which are described below. Other brainstem and hypothalamic neurons also project to the intermediolateral nucleus in the

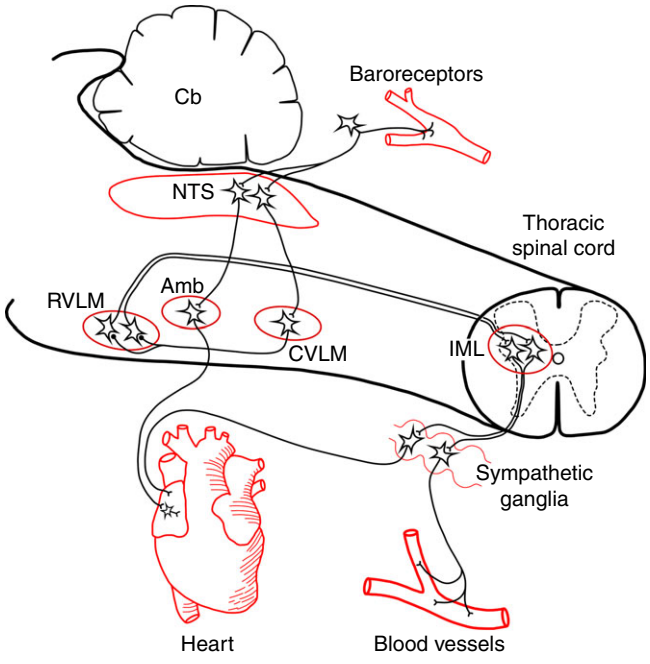


FIGURE 35.9 Anatomical schematic of the organization of the arterial baroreceptor reflex pathways providing short-term stabilization of arterial pressure through feedback control of cardiac and vascular performance (see text for description). Amb, nucleus ambiguus, containing cardiac vagal preganglionic neurons; Cb, cerebellum; CVLM, caudal ventrolateral medulla, containing GABAergic, inhibitory interneurons; IML, intermediolateral nucleus of the thoracic spinal cord, containing cardiac and vasoconstrictor sympathetic preganglionic neurons; NTS, nucleus of the tractus solitarius, containing second-order, barosensory neurons; RVLM, rostral ventrolateral medulla, containing sympathetic premotor neurons projecting to the spinal cord to control the heart and blood vessels.

thoracic spinal cord and thus also act as sympathetic premotor neurons. However, interruption of their activity does not eliminate basal sympathetic tone.

How and where is basal sympathetic tone generated? Understanding the generation and regulation of the basal sympathetic tone to the cardiovascular system is highly relevant, since exaggerated sympathetic outflow to the vasculature contributes to the elevated blood pressure in several forms of hypertension and an enhanced sympathetic outflow to the heart is a precipitating factor in cardiac arrhythmias leading to sudden cardiac death (Box 35.4). Blockade of various RVLM inputs has not been fruitful in identifying the primary sources controlling their activity, a key mystery in explaining the generation of the tonic cardiovascular sympathetic nerve activity. Although the RVLM neuronal pacemaker-like discharge seen under certain *in vitro* conditions was once proposed to provide the tonic drive, this does not appear to be the case because, under normal conditions, their rhythmic discharge is not endogenous, but rather is due to synaptic drive. Thus, although the RVLM sympathetic premotor neurons are essential elements in the circuit generating the basal cardiovascular sympathetic outflow (Fig. 35.8), the network or cellular mechanisms responsible for their basal activity remain unknown.

The parasympathetic (vagal) preganglionic neurons influencing cardiac function are strongly excited by inputs from the NTS which has neurons receiving input from arterial baroreceptors that sense pressure in the major arteries, thereby providing the afferent drive for vagal baroreceptor reflex bradycardia (Fig. 35.9). Although the central circuits underlying the parasympathetic vasodilatory control of certain vascular beds (see above) remain relatively unknown, this parasympathetically mediated redistribution of blood flow is critical for the successful performance of certain physiological functions, such as reproduction.

Summary

Neural circuits, primarily within the brainstem, are responsible for generating the modest level of vasoconstrictor sympathetic nerve activity to blood vessels that must be sustained to provide the vascular tone necessary to maintain a sufficient level of arterial pressure to perfuse all the capillary beds in the body, but especially those in the brain and heart, with the O_2- and nutrient-rich blood necessary for survival. Although the cellular and neural network mechanisms through which this basal sympathetic nerve activity is generated remain unknown, the vasoconstrictor sympathetic premotor neurons in the RVLM are the key medullary output neurons of this generating network, providing the glutamatergic excitation necessary to drive the spinal sympathetic circuits whose activity maintains vascular tone. Similarly organized, but distinct brainstem and spinal circuits generate and control the cardiac and other sympathetic outflows, contributing to the optimized tissue function necessary for homeostasis, the support of behavior and organismal survival.

BOX 35.4

CARDIOVASCULAR DISEASE

Alterations in the autonomic neural regulation of the cardiovascular system characterize the major cardiovascular diseases—contributing significantly to the etiology of the disease and to exacerbation of the symptoms and morbidity. Elevated sympathetic vasoconstrictor outflow plays a role in the increased vascular resistance seen in essential (primary) hypertension, the most prevalent form of high blood pressure and a major risk factor for stroke and congestive heart failure. Although the cause or trigger for this augmented stimulation of the CNS circuits generating vasoconstrictor sympathetic nerve activity is unknown, elevated dietary salt or increased central production of the hormone angiotensin II, both of which increase the discharge of RVLM sympathetic premotor neurons, are thought to contribute. Following myocardial infarction (heart attack), the cardiac sympathetic outflow is increased, at least partially the result of reflex drive to compensate for the reduced cardiac output, and this sympathetic drive to increase heart rate and cardiac contractility increases the stress on the weakened cardiac muscle, contributing to congestive heart failure and predisposing to arrhythmias. An imbalance in the neural regulation of the heart beat, with exaggerated cardiac sympathetic nerve activity and reduced cardiac vagal outflow, is a main cause of lethal arrhythmias and sudden cardiac death. Attempts to reduce mortality in such patients include surgical removal of the right stellate ganglion which contains the sympathetic ganglion neurons controlling heart rate. Increased understanding of the neural circuits and neurotransmitters controlling and generating sympathetic nerve activity to the cardiovascular system holds the promise of improved therapeutic approaches to a variety of cardiovascular diseases.

Shaun F. Morrison

SENSORY REGULATION OF RESPIRATORY AND CARDIOVASCULAR SYSTEMS

The principal homeostatic goal in the control of breathing is the regulation of the pressures of O_2 and CO_2 and pH in arterial blood—that is, arterial PO_2 = 80–100 mmHg, PCO_2 = 35–45 mm Hg and pH = 7.35–7.45. Such regulation requires their constant measurement by sensors connected in a feedback network to control breathing. Decreases in either PO_2 or pH or an increase in PCO_2 stimulate breathing, whereas increases in PO_2 or pH or a decrease in PCO_2 depress breathing. Similarly, a principal homeostatic goal in the regulation of the cardiovascular system is the maintenance of adequate blood flow to the brain. To meet this requirement, the CNS receives baroreceptor sensory input related to the stretch in the walls of the major arteries and modifies cardiovascular performance to sustain an arterial pressure sufficient to perfuse the brain with oxygenated blood.

Controlling Breathing and Cardiovascular Function; O_2 and CO_2 Are Measured By Chemoreceptors

The consequences of anoxia are catastrophic (particularly for the brain), but the effects of modestly depressed or elevated PO_2 are benign (due in part to the unique way blood hemoglobin reversibly binds O_2). However, small changes in PCO_2 (with consequent changes in pH) can profoundly affect cellular metabolism. Ventilation is very sensitive to small changes in PCO_2: a 1 mmHg increase in PCO_2 leads to a 2 liter × min^{-1} increase in ventilation at rest (~5 liter × min^{-1} in an adult human). In other words, a 2.5% increase in PCO_2 at rest leads to a 40% increase in ventilation! Thus, at rest and typically in exercise, ventilation is controlled mostly to regulate CO_2; however, in severe hypoxia, everything else is ignored. For example, compared to the normal ambient PO_2 at sea level (~160 mmHg), the exceptionally low PO_2 (~40 mm Hg) at the top of Mount Everest results in such a strong ventilatory drive that PCO_2 in climbers without supplemental O_2 is exceptionally low (~10 mm Hg compared with normal ~40 mm Hg). The sites and mechanisms of CO_2 chemoreception are the subject of intense scrutiny (Feldman, Mitchell, & Nattie, 2003). While the carotid bodies are sensitive to changes in PCO_2, a robust CO_2 response remains in peripherally chemo-denervated, vagotomized mammals, indicating the existence of intracranial sensors for CO_2 or related variables (pH, HCO_3^-). A number of structures are implicated in intracranial chemotransduction (Feldman et al., 2003; Huckstepp & Dale, 2011). Among these, neurons within the RTN (Guyenet et al., 2010) and in the raphe (Richerson, 2004; Ray et al., 2011) respond robustly to changes in CO_2/pH within the physiological range and project to respiratory nuclei.

The mammalian brain is extremely sensitive to O_2 deprivation: several minutes of anoxia can initiate a cascade leading to loss of consciousness and neuronal death. Accordingly, the arterial chemoreceptors, the principal O_2 sensors for the entire body, are located in the carotid bodies in the bifurcation of the common carotid artery, the main portal through which O_2-enriched blood is transported to the brain. Arterial chemoreceptor afferents enter the brain via the glossopharyngeal nerve (cranial nerve IX) and terminate on second-order sensory neurons located primarily in the commissural region of the NTS.

Arterial chemoreceptor signals account for only a small part of the chemical drive to breathe during wakefulness, as breathing 100% O_2 reduces peripheral chemoreceptor discharge to near zero but only decreases ventilation by ~15% in awake mammals. Decreasing PO_2 has relatively little influence on chemoreceptor activity and ventilation until it falls to ≤60 mm Hg. With further reductions in PO_2, chemoreceptor discharge and ventilation increase exponentially.

In contrast to the global effects of hypoxia on breathing, the cardiovascular response is compartmentalized, directed toward conserving the O_2 availability for the brain by increasing brain blood flow while simultaneously reducing blood flow to other tissues. This is accomplished primarily by a widespread sympathetically induced vasoconstriction, which is driven by the arterial chemoreceptors and mediated by a strong activation of RVLM sympathetic premotor neurons. Although the brainstem circuitry underlying arterial chemoreceptor-stimulated excitation of RVLM sympathetic premotor neurons remains to be identified, a role for the RTN has been strongly implicated (Moreira, Takakura, Colombari, & Guyenet, 2006). In extreme cases, vasoconstriction evoked by severe hypoxia can reduce blood flow sufficiently to cause organ damage—for example, renal failure. Interestingly, although diving mammals such as seals also conserve brain O_2 availability by slowing their heart rate (by vagal activation) and maximally constricting their muscle blood vessels during their dives (hypoxic breath-holding), they avoid hypoxic tissue damage and maintain use of skeletal muscle through adaptations that include a prominent increase in myoglobin concentration.

Baroreceptor Reflex Regulation of Cardiovascular Performance

With the primary goal of maintaining adequate perfusion of the brain with sufficient O_2-rich arterial blood to support its continuous and high level metabolism, two regulatory circuits are critical for the maintenance of blood pressure (via the baroreceptor reflex) and for the distribution of blood flow (via the arterial chemoreceptor

reflex). The baroreceptor reflex can be modeled as a simple negative feedback circuit in which changes in arterial pressure evoke compensatory changes in cardiovascular sympathetic nerve activity and cardiac parasympathetic vagal nerve activity to affect cardiac output and vascular resistance in an effort to return arterial pressure toward its baseline level. Figure 35.9 illustrates the pathways mediating the baroreceptor reflex-mediated changes in autonomic outflows. The sensors in the baroreceptor reflex are stretch receptors (baroreceptors), sensitive to changes in the pressure across the walls of the major arteries, and with axons in the vagus and glossopharyngeal nerves that project to second-order NTS neurons (Fig. 35.9). Thus, the activity of NTS barosensory neurons tracks increases and decreases in arterial pressure, including the rapid rise and fall in arterial pressure that occurs during the systolic phase of each cardiac cycle. Second-order glutamatergic barosensory neurons in the NTS project to GABAergic neurons in the caudal ventrolateral medulla (CVLM) that, in turn, project to the RVLM to inhibit cardiovascular sympathetic premotor neurons (Fig. 35.9). Thus, for example, a fall in arterial pressure will induce the following sequence: decreased activity of barosensory NTS neurons leading to disfacilitation of CVLM inhibitory interneurons and subsequent disinhibition of RVLM sympathetic premotor neurons causing increased preganglionic activity resulting in cardiac stimulation and vasoconstriction, acting together to counter the fall in arterial pressure. Concurrently, the reduced firing of NTS barosensory neurons decreases the excitation of cardiac vagal preganglionic neurons in the nucleus ambiguus (Fig. 35.9), thereby diminishing their inhibitory influence on the heart, effectively amplifying the effects of increases in cardiac sympathetic nerve activity. Each of the synaptic integration sites in the baroreceptor reflex pathway present the potential for other inputs to modulate the effectiveness of the reflex. For instance, both arterial pressure and heart rate must rise to provide adequate perfusion of skeletal muscle during escape behavior (or exercise, see below) and this is accomplished, in part, by presynaptic inhibition of the baroreceptor afferent inputs to second-order barosensory NTS neurons from hypothalamic neurons activated during these behaviors.

Summary

Chemoreceptors in the carotid bodies and in the brain provide sensory input to the central circuits controlling breathing and cardiovascular function. PCO_2 is normally regulated within narrow limits over a wide range of metabolic demands and this usually assures an adequate O_2 supply to the brain. However, during significant hypoxia, arterial chemoreceptors sensitive to O_2 can provide a powerful stimulus to increase breathing. Central chemoreceptors are exquisitely sensitive to changes in PCO_2 and in pH and can evoke large and rapid increases in respiration. Baroreceptors transduce the pressure in the major arteries to neural signals that modulate the autonomic outflows to the cardiovascular system, thereby stabilizing arterial pressure and maintaining adequate blood flow to the brain.

MODULATION AND PLASTICITY OF RESPIRATORY MOTOR OUTPUT

In meeting the homeostatic goal of maintaining normal arterial PO_2 and PCO_2, breathing must continually be adapted to meet changes in the state of the body—for example, posture—and in the environment (e.g., sojourning to or living at high altitude), and, in particular, to compensate for a variety of chronic cardiopulmonary diseases that result in low arterial PO_2.

Adaptations in the control of breathing in response to hypoxemia are examples of neural plasticity (Feldman et al., 2003). Inhaling an hypoxic gas mixture for 10–20 min elicits a progressive increase in ventilation. This slow increase in breathing at the onset of hypoxia is matched with a slow posthypoxic decline back to normal ventilation that is termed *short-term potentiation* (arrowheads in Fig. 35.10). Thus, even though the afferent input to the respiratory system from arterial and central chemoreceptors can change abruptly, the respiratory response consists of slower, smoother changes. This is important since abrupt changes can result in an unstable breathing pattern such as Cheyne-Stokes respiration (Fig. 35.4 and see Box 35.2), often seen in patients with severe congestive heart failure.

A longer-lasting plasticity is revealed following a repetitive series of brief episodes of hypoxia, as may occur in a patient with sleep apnea (see Box 35.5). As illustrated in Figure 35.10, following the final hypoxic episode, ventilation transiently returns toward control levels, but is followed by a progressive augmentation that lasts for several hours termed *respiratory long-term facilitation.* Like other forms of neural plasticity in the brain, respiratory long-term facilitation is dependent on activation of serotonergic and noradrenergic inputs to respiratory motoneurons. This activates protein kinases, such as protein kinase C, thereby initiating an intracellular signaling cascade that amplifies the synaptic currents arising from glutamatergic respiratory premotoneurons. In response to a single prolonged hypoxic exposure, protein phosphatases constrain activity in this pathway and prevent long-term facilitation. However, repetitive bouts of hypoxia decrease the activity of phosphatases such as the serine/threonine phosphatases, thereby removing this constraint and permitting the development of long-term facilitation (Wilkerson, Macfarlane, Hoffman, & Mitchell, 2007).

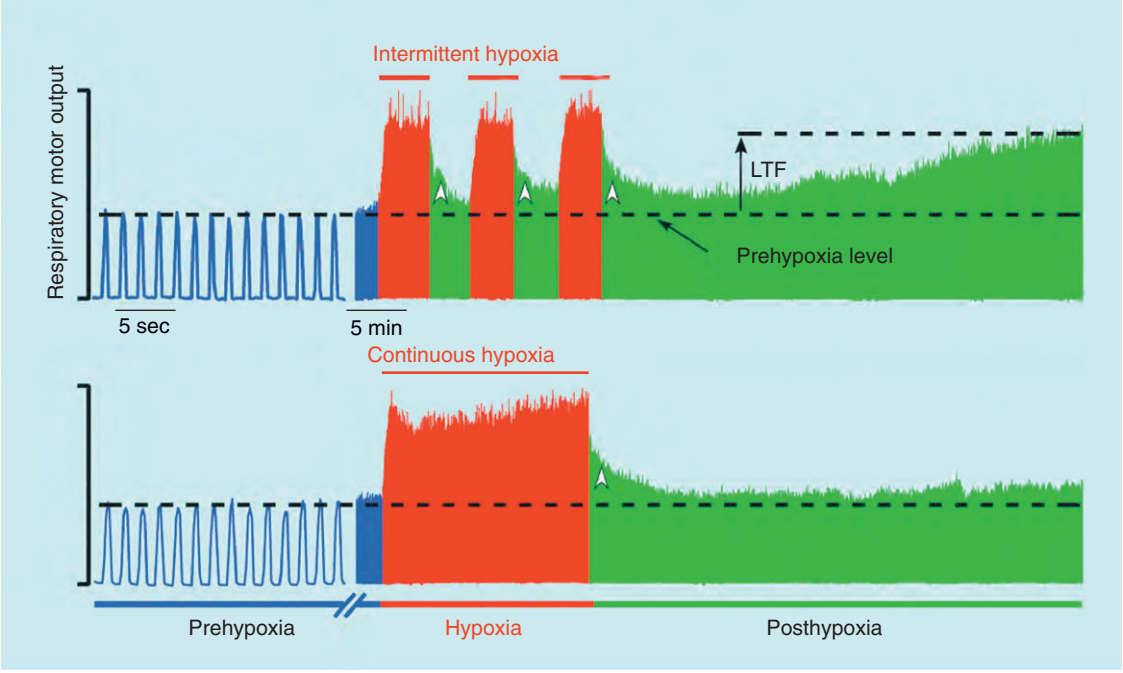


FIGURE 35.10 Respiration-related neural activity in an anesthetized and artificially ventilated rat exhibits *respiratory long-term facilitation*. Peak electrical activity in the phrenic nerve corresponds to the magnitude of inspiration. When arterial O_2 and CO_2 are normal (prehypoxia), the phrenic bursts are rhythmic and consistent (shown with expanded time scale). Upon exposure to intermittent (upper trace) or continuous (lower trace) hypoxia, phrenic bursts increase in amplitude, reflecting the drive to take deeper breaths when hypoxia-sensitive chemoreceptors are activated. In both patterns of hypoxia, phrenic nerve activity slowly returns nearly to baseline over a period of several minutes in a process termed short-term potentiation (STP; open arrowheads). Following intermittent, but not continuous hypoxia, phrenic burst amplitude progressively increases over an hour. This slow increase reflects respiratory long-term facilitation (LTF). *Courtesy of T. L. Baker and G. S. Mitchell.*

BOX 35.5

DISORDERS OF BREATHING ARE WIDESPREAD AND SERIOUS

Breathing at rest is normally controlled to maintain arterial PCO_2 at 40±2 mm Hg. However, cardiopulmonary pathology can increase PCO_2 markedly. Narcotics, barbiturates, and most general anesthetics depress breathing and at high doses, as can happen when self-administered, these drugs can cause death by respiratory failure. For opioids, respiratory depression appears to be due to depressant actions directly on preBötzinger complex neurons.

Sleep apnea. Sleep is associated with a modest decrease in ventilation and a reduced responsiveness to deviations in PCO_2 and PO_2. With the onset of sleep, the breathing pattern can become unstable, and apnea (defined as at least 10 s without breathing) can occur. In the most common form, *obstructive sleep apnea*, there is a loss of tone in airway muscles and tongue, so that the reduction in airway pressure during inspiration can pull in on the walls of the upper airway sufficiently to cause vibration—that is, snoring, even airway obstruction. Obstruction reduces or abolishes ventilation, raising CO_2 and lowering O_2. This in turn can cause arousal, and with it restoration of airway muscle tone and airway patency. In serious cases, the cycle of sleep, airway obstruction, and hypoxia-induced waking repeats hundreds of times every night. The marked disruption of sleep can cause debilitating hypersomnolesence, and severe obstructive sleep apnea can also have other sequelae, such as hypertension. In *central sleep apnea*, pauses in breathing result from the failure of respiratory drive. Mechanisms underlying central apneas are not understood, but the problem is prevalent during sleep at the end stages of a variety of neurodegenerative diseases, including amyotrophic lateral sclerosis, Parkinson's disease, and multiple systems atrophy (where there is evidence of loss of preBötC neurons).

Sudden infant death syndrome (SIDS). SIDS is defined as the unexpected sudden death of an infant or young child that cannot be explained by a postmortem examination. SIDS is the leading cause of death of infants between 1 month and 1 year of age in the United States. Although a

BOX 35.5 *(cont'd)*

number of etiologies are likely to contribute to SIDS, current hypotheses tend to focus on abnormalities of cardiorespiratory control. The apnea hypothesis of SIDS attributes apnea to disorders of both the chemical control of breathing and the arousal response to insufficient ventilation. Infants who later succumb to SIDS have been described as exhibiting irregular breathing patterns, including periods of apnea, and depressed arousal responses to hypoxia or hypercapnia. Epidemiologically, SIDS has been associated with infants sleeping face down, where pillows, blankets, and mattress can limit the diffusion of expired air, resulting in elevated CO_2 and lower O_2 in inspired air and, consequently, in arterial blood. This in turn could depress breathing sufficiently to cause respiratory arrest and death.

Genetic Diseases Affecting Breathing

Congenital Central Hypoventilation Syndrome (CCHS; Weese-Mayer et al., 2009). CCHS is a potentially life-threatening genetic disorder typically presenting in newborns and characterized by respiratory insufficiency especially during sleep. This breathing phenotype is the consequence of an impaired ventilatory response to hypercapnia (increased blood PCO_2) and hypoxemia (decreased blood PO_2). Additionally, the disease is often associated with functional impairments of the autonomic nervous system and tumors of neural crest derivatives. The defining mutation in CCHS is a polyalanine expansion in *Phox2b*. The extent of the polyalanine repeat mutation correlates with the severity of the disease, with the most severe cases requiring continuous artificial ventilation. The offspring of mice genetically modified to contain the polyalanine expansion have irregular breathing at birth, do not increase their breathing in response to CO_2 stimulation of chemoreceptors, and die shortly after birth. Postmortem examination of these mice provided a likely explanation for the depression of breathing. Chemoreceptor neurons in the RTN (Fig. 35.2) normally express Phox2b but in mouse pups with the Phox2b polyalanine expansion these neurons were absent. Thus, the *Phox2b* mutation appears to result in a loss of RTN neurons that reduces the normal respiratory drive arising from chemoreceptors.

Rett Syndrome (RTT; Lioy, Wu, & Bissonnette, 2011). RTT is a developmental disorder of the central nervous system resulting from mutations in the X-linked gene methyl-CpG-binding protein 2 (*Mecp2*) and is usually seen in girls. It results in severe neurological symptoms that include speech and movement disorders as well as irregular breathing (e.g., Cheyne-Stokes breathing; Fig 35.4 and Box 35.2) that can include life-threatening periods of apnea. The onset of symptoms typically begins at about 6–18 months of age at which time there is a slowing down of motor skills and then a regression in communication skills.

Jack L. Feldman and Donald R. McCrimmon

Exercise: An Example of Cardiopulmonary Integration

Physical activity involving repetitive movement of large muscles significantly increases the metabolic consumption of O_2 and production of CO_2. Even during mild activity, such as walking, the metabolic changes require substantial ventilatory and cardiac output increases to maintain arterial and brain levels of CO_2 and O_2 within normal limits. Impressively, the cardiorespiratory changes are proportional to the changes in metabolism and there is almost no change in arterial PCO_2 or PO_2 with mild or moderate intensity exercise. During high intensity exercise, the increase in breathing can result in a decrease (!) in arterial PCO_2. The primary cardiovascular adjustments accompanying exercise include not only the sympathetically mediated increases in heart rate and cardiac contractility necessary to elevate cardiac output required to support increased muscle metabolism, but also a widespread sympathetic vasoconstriction. This increased sympathetic input to the vasculature plays a significant role in optimizing exercise performance. Blood volume "stored" in the veins at rest is brought into the arterial circulation to increase the capacity for O_2 transport to exercising muscle. Arterial blood pressure is maintained in the face of a marked reduction in total peripheral vascular resistance due to the substantial vasodilation occurring in exercising muscle (driven mostly by local metabolic signals—for example, adenosine—rather than by a selective inhibition of sympathetic tone to exercising muscle). Metabolism and O_2 consumption in nonexercising tissues is significantly reduced—for example, the reduced blood flow to the GI tract accounts for the cramping that occurs with intense exercise immediately after a large meal. Sympathetically driven adrenal epinephrine secretion further supports both the increased cardiac output and the skeletal muscle hyperemia. Because the large amount of heat produced by exercising muscles (and the respiratory muscles engaged to support exercise) must be eliminated to prevent a significant rise in body temperature (potentially lethal hyperthermia), secondary, but critically important cardiovascular changes are engaged as exercise

progresses: a reduction in sympathetic vasoconstrictor tone to the skin increases skin blood flow and heat loss, and sympathetic vasoconstriction to the kidney is increased to retain water and electrolytes that are required to maintain blood volume during the increased sweat production (as much as 1.5 l/h) required to enhance heat loss from the skin.

Despite its utility as a model of neural integration and widespread scientific fascination with the topic, the physiological mechanisms underlying the matching of changes in cardiorespiratory function to exercise remain mysterious. Since there is normally little change in CO_2 or O_2, classic negative feedback from arterial or brain chemoreceptors cannot adequately account for these cardiorespiratory adjustments. Thus, feedforward, descending command from higher brain systems and sensory input from joint and muscle receptors appears to be important.

The volitional initiation of exercise, signaled by suprapontine structures, likely involves activation of spinal pathways for locomotion and concurrent feedforward activation of brainstem mechanisms controlling breathing and cardiovascular function. This allows pulmonary ventilation and gas transport to change appropriately and quickly before there are any significant changes in blood gases (as would occur if the mechanism was strictly based on feedback). Commands for the activation of pulmonary and locomotor muscles presumably arise in similar ways in the motor cortex, thalamus, and basal ganglia. Once the motor command has been initiated, respiratory and locomotor activities are likely coordinated at subcortical levels, including the hypothalamus and periaqueductal gray. Electrical or chemical activation of the hypothalamus in the anesthetized, paralyzed cat elicits fictive locomotion paired with proportional increases in respiratory motor output (suggestive of exercise-related increases in ventilation, termed exercise hyperpnea) and redistribution of blood flow consistent with the induction of locomotion (Waldrop, Eldridge, Iwamoto, & Mitchell, 1996). Because the animals were paralyzed, muscle activation did not occur, and thus the changes in breathing were the result of feedforward control rather than sensory feedback from joint or muscle receptors.

At the level of the brainstem, coordination of the respiratory and cardiovascular responses likely involves circuitry in the dorsolateral pons, in the vicinity of the PRG. Descending inputs from the hypothalamus, periaqueductal gray, and dorsolateral pons activate medullary respiratory circuits to increase both the depth of each breath—that is, tidal volume—and respiratory frequency, which almost certainly involves modulation of the activity of the rhythm generating circuitry in the preBötC. The need for an increase in the size of each breath means there is also a parallel increase in the activity of pattern forming circuits in, for example, the rVRG. There is also an increase in the contribution of expiratory muscles. Expiratory rhythm generating neurons in the region of the pFRG appear to be synaptically inhibited at rest and this inhibition may be removed by signals related to exercise, possibly through the influence of descending inputs from the higher brain structures. Through reciprocal connections between the pFRG and preBötC, activation of the pFRG may contribute to the increase in breathing frequency. Activation of the pFRG will also increase tidal volume through activation of expiratory pattern-forming circuits in the caudal VRG.

The increases in cardiac, vasoconstrictor, and adrenal sympathetic outflows necessary for exercise are mediated by enhanced discharge of the relevant sympathetic premotor neurons in the RVLM. Although the exercise-related stimulus and the source of the increased excitatory drive to these premotor neurons are unknown, their excitatory inputs from central respiratory neurons are likely to contribute significantly. Also contributing to the sympathetic activation is a centrally-mediated reduction in the effectiveness of the inhibitory baroreceptor reflex, due to modulation of the synaptic efficacy in the NTS, which would otherwise inhibit sympathetic outflow as arterial blood pressure increases during exercise. The increased central respiratory drive may also contribute to the activation of adrenal sympathetic premotor neurons in the RVLM during exercise, which leads to secretion of adrenal epinephrine and a potent stimulation of cardiac output and skeletal muscle vasodilation. As the body heat load rises during exercise, warm-sensitive, thermoregulatory neurons in the preoptic hypothalamus increase their inhibitory influence on the cutaneous vasoconstrictor premotor neurons in the raphe nuclei of the ventromedial medulla to dilate skin blood vessels. Simultaneously, these preoptic neurons excite the sympathetic outflow to sweat glands to augment heat loss. As the osmolarity of the blood increases due to sweating, a specialized, osmosensitive neural network in the anterior hypothalamus increases the activity of renal sympathetic premotor neurons in the RVLM to augment the sympathetic outflow to the kidney to reduce renal filtration and help sustain blood volume.

Summary

Breathing and cardiovascular function are fundamental and intertwined physiologic functions that must be controlled vigilantly by the brain. This chapter has explored the basic mechanisms underlying rhythm generation, sensory processing, and motor output for these regulatory systems.

References

Adrian, E. D., & Buytendijk, F. J. J. (1931). Potential changes in the isolated brainstem of goldfish. *Journal of Physiology (London), 71,* 121–135.

Alheid, G. F., & McCrimmon, D. R. (2008). The chemical neuroanatomy of breathing. *Respiratory Physiology & Neurobiology, 164,* 3–11.

Canning, B. J., & Spina, D. (2009). Sensory nerves and airway irritability. *Handbook of Experimental Pharmacology, 194,* 139–183.

Feldman, J. L. (1986). Neurophysiology of breathing in mammals. In F. E. Bloom (Ed.), *Handbook of physiology; section I: the nervous system; volume IV: intrinsic regulatory systems of the brain* (pp. 463–524). Bethesda, MD: American Physiological Society.

Feldman, J. L., & Del Negro, C. A. (2006). Looking for inspiration: new perspectives on the generation of respiratory rhythm. *Nature Reviews Neuroscience, 7,* 232–242.

Feldman, J. L., Mitchell, G. S., & Nattie, E. E. (2003). Breathing: Rhythmicity, plasticity, chemosensitivity. *Annual Review of Neuroscience, 26,* 239–266.

Gesell, R., Bricker, J., & Magee, C. (1936). Structural and functional organization of the central mechanism controlling breathing. *The American Journal of Physiology, 117,* 423–452.

Guyenet, P. G., Stornetta, R. L., Abbott, S. B., Depuy, S. D., Fortuna, M. G., & Kanbar, R. (2010). Central CO2 chemoreception and integrated neural mechanisms of cardiovascular and respiratory control. *Journal of Applied Physiology, 108*(4), 995–1002.

Hayashi, F., & McCrimmon, D. R. (1996). Respiratory motor responses to cranial nerve afferent stimulation in rats. *The American Journal of Physiology, 271*(4 Pt 2), R1054–1062.

Huckstepp, R. T., & Dale, N. (2011). Redefining the components of central CO2 chemosensitivity - towards a better understanding of mechanism. *Journal of Physiology, 589*(Pt 23), 5561–5579.

Janczewski, W. A., & Feldman, J. L. (2006). Distinct rhythm generators for inspiration and expiration in the juvenile rat. *Journal of Physiology, 570*(Pt 2), 407–420.

Kubin, L., Alheid, G. F., Zuperku, E. J., & McCrimmon, D. R. (2006). Central pathways of pulmonary and lower airway vagal afferents. *Journal of Applied Physiology, 101*(2), 618–627.

Legallois, M. (1813). *Experiments on the principle of life*. Philadelphia, PA: Thomas.

Lioy, D. T., Wu, W. W., & Bissonnette, J. M. (2011). Autonomic dysfunction with mutations in the gene that encodes methyl-CpG-binding protein 2: insights into Rett syndrome. *Autonomic Neuroscience: Basic & Clinical, 161*(1–2), 55–62.

Lorry, M. (1760). Les mouvements du cerveau. *Mem. Math. Phys. Pres. Acad. Roy. Sci. Div. Sav. Paris, 3,* 344–377.

Mellen, N. M., Janczewski, W. A., Bocchiaro, C. M., & Feldman, J. L. (2003). Opioid-induced quantal slowing reveals dual networks for respiratory rhythm generation. *Neuron, 37*(5), 821–826.

Moreira, T. S., Takakura, A. C., Colombari, E., & Guyenet, P. G. (2006). Central chemoreceptors and sympathetic vasomotor outflow. *The Journal of Physiology, 577*(Pt 1), 369–386.

Pagliardini, S., Janczewski, W. A., Tan, W., Dickson, C. T., Deisseroth, K., & Feldman, J. L. (2011). Active expiration induced by excitation of ventral medulla in adult anesthetized rats. *Journal of Neuroscience, 31,* 2895–2905.

Plum, F., & Posner, J. B. (1980). *The diagnosis of stupor and coma* (3rd ed.). Contemporary Neurology Series Vol. 19. Philadelphia, PA: Davis.

Ramon y Cajal, S. (1909). *Histologie du Systeme Nerveux de l'homme et des Vertebres*. Paris: Maloine.

Ray, R. S., Corcoran, A. E., Brust, R. D., Kim, J. C., Richerson, G. B., Nattie, E., & Dymecki, S. M. (2011). Impaired respiratory and body temperature control upon acute serotonergic neuron inhibition. *Science, 333*(6042), 637–642.

Rekling, J. C., & Feldman, J. L. (1997). Calcium-dependent plateau potentials in rostral ambiguus neurons in the newborn mouse brain stem in vitro. *Journal of Neurophysiology, 78*(5), 2483–2492.

Richerson, G. B. (2004). Serotonergic neurons as carbon dioxide sensors that maintain pH homeostasis. *Nature Reviews Neuroscience, 5*(6), 449–461.

Smith, J. C., Ellenberger, H. H., Ballanyi, K., Richter, D. W., & Feldman, J. L. (1991). Pre-Bötzinger complex: a brainstem region that may generate respiratory rhythm in mammals. *Science, 254*(5032), 726–729.

Suzue, T. (1984). Respiratory rhythm generation in the in vitro brain stem-spinal cord preparation of the neonatal rat. *The Journal of Physiology, 354,* 173–183.

Tan, W., Janczewski, W. A., Yang, P., Shao, X. M., Callaway, E. M., & Feldman, J. L. (2008). Silencing preBötzinger complex somatostatin-expressing neurons induces persistent apnea in awake rat. *Nature Neuroscience, 11*(5), 538–540.

Weese-Mayer, D. E., Rand, C. M., Berry-Kravis, E. M., Jennings, L. J., Loghmanee, D. A., Patwari, P. P., & Ceccherini, I. (2009). Congenital central hypoventilation syndrome from past to future: model for translational and transitional autonomic medicine. *Pediatric Pulmonology, 44*(6), 521–535.

Wilkerson, J. E., Macfarlane, P. M., Hoffman, M. S., & Mitchell, G. S. (2007). Respiratory plasticity following intermittent hypoxia: roles of protein phosphatases and reactive oxygen species. *Biochemical Society Transactions, 35*(Pt 5), 1269–1272.

Suggested Readings

Guyenet, P. G. (2006). The sympathetic control of blood pressure. *Nature Reviews Neuroscience, 7,* 335–346.

Malpas, S. C. (2010). Sympathetic nervous system overactivity and its role in the development of cardiovascular disease. *Physiological Reviews, 90,* 513–557.

Waldrop, T. G., Eldridge, F. L., Iwamoto, G. A., & Mitchell, J. H. (1996). Central neural control of respiration and circulation during exercise. In L. B. Rowell & J. T. Shepherd (Eds.), *Handbook of physiology sec 12 exercise: regulation and integration of multiple systems* (pp. 333–380). New York: Oxford University Press.

J.L. Feldman, D.R. McCrimmon, and S.F. Morrison

CHAPTER

36

Food Intake and Metabolism

Eating is a familiar behavior. In humans, it is strongly influenced by cultural, social, and experiential factors. Consequently, eating has been a subject of considerable interest to social and behavioral scientists. In mammals, including humans, food intake is also a regulatory behavior with the primary function of supporting the continuous energy demands of body tissues. When considered from this biological perspective, food intake is influenced by hunger, satiety, and the physiological mechanisms that couple eating with internal caloric supplies and a stable body weight. This chapter discusses the signals important for the central nervous system's control of food intake and glucose homeostasis, and also describes mechanisms thought to integrate those signals.

Despite decades of investigation, considerable differences of opinion still exist about how food intake is controlled. The multiple views follow one of two principles. The first considers eating to be a consequence of depleted energy stores in adipose tissue, reduced use of energy-rich *metabolic fuel* (glucose or lipid) in some critical tissue, or both. In this schema, the purpose of eating is to restore energy reserves in adipose tissue or to increase fuel utilization to some desired level, thereby eliminating the signal to eat. This traditional negative feedback model is not the approach taken in this chapter. Rather, we subscribe to a second principle, in which hunger is thought to be initiated in part by the gradual disappearance of inhibitory satiety stimuli that were generated by the previous meal. In addition, meal initiation may occur based on learning, habits, and various social and other situations, which allow animals to take maximal advantage of their environment. According to this view, *caloric homeostasis* influences meals, albeit indirectly, and eating contributes to caloric homeostasis by providing nutrients. The storage and use of those nutrients are regulated independently by physiological mechanisms described later. We describe the properties of caloric homeostasis before we consider the control of food intake.

CALORIC HOMEOSTASIS

Homeostatic Mechanisms Provide a Continuous Supply of Metabolic Fuels to Cells

The purpose of caloric homeostasis is to preserve cellular metabolism. Cells oxidize metabolic fuels to drive all cellular processes; thus, the higher the cellular activity, the greater the demand for energy. Because most cells have limited amounts of stored energy, they rely on a steady supply of calories and oxygen from the bloodstream. Oxygen is dependably and instantaneously available via the respiratory system and is not stored in the body. In contrast, food calories can be scarce and require time after ingestion to become available to cells in significant quantities. One consequence of this functional organization is the capacity of animals to store sufficient energy to bridge long intervals during which no food is eaten.

Three categories of macronutrients—carbohydrates, lipids, and proteins—provide usable energy, but the use of specific macronutrients by the body varies depending on the tissue. Most tissues can oxidize carbohydrates in the form of *glucose* or lipids in the form of *free fatty acids*, depending on the availability of these nutrients and the levels of certain hormones in the blood. Notable exceptions are the liver, which requires lipids for proper functioning, and the brain, which has a large, continuous need for glucose despite the ability to oxidize lipids in the form of *ketone bodies*. When the supply of glucose to the brain is compromised, neurons cease to function and in a few minutes consciousness is lost. Death will ensue unless glucose delivery to the

© 2013 Elsevier Inc. All rights reserved.

brain is restored. Therefore, maintenance of circulating glucose in sufficient amounts to support normal brain function is a critical goal of caloric homeostasis.

Two distinct metabolic states are defined by the availability of recently consumed food to cells. The *prandial*, or fed, state is characterized by an abundance of newly ingested and absorbed nutrients in the blood. These molecules are sequestered rapidly in tissues to prevent them from being excreted wastefully in urine. The *postabsorptive*, or fasted, state is characterized by the absence of calories entering the circulation from the gastrointestinal tract and a consequent reliance on energy from metabolic fuels less recently consumed and stored. These stores are released gradually into the blood during a fast. In both states, tissues take nutrients from the blood as needed for cellular metabolism.

Many tissues store carbohydrates in the form of *glycogen*, a polymer of glucose; the liver and skeletal muscles have the largest depots. Energy is stored more efficiently, mainly in adipose tissue, as *triglyceride*, each molecule of which consists of glycerol with three attached fatty acids. During the prandial period, newly ingested food is used immediately by the body or is stored as glycogen or triglyceride. Excess carbohydrate is largely converted to lipid (*lipogenesis*) because glycogen storage capacity is limited. During the fasting period, liver glycogen is converted back to glucose (*glycogenolysis*), which enters the blood and is available to all tissues. Similarly, stored triglycerides are mobilized from adipose tissue (*lipolysis*) and enter the circulation as fatty acids and glycerol. The fatty acids are used by tissues as needed or are converted to ketone bodies (*ketogenesis*), whereas glycerol is converted to glucose (*gluconeogenesis*) (Fig. 36.1).

The liver is the key organ of energy traffic. Lipogenesis (which also occurs in adipose tissue) and glycogen formation (which also occurs in skeletal muscle) occur in the liver during the prandial period, and glycogenolysis, ketogenesis, and gluconeogenesis occur during the fasting period. These processes are regulated (discussed later), as are delivery of metabolic fuels (from the small intestine into the circulation), storage of excess fuels, and mobilization of stored energy. The control system involves interplay among several hormones and the sympathetic and parasympathetic divisions of the autonomic nervous system.

Insulin is the Key Hormone Affecting Caloric Homeostasis

Secretion of the peptide hormone *insulin* from B cells of the pancreatic islets is influenced by several factors, among which interstitial glucose is critical. Insulin secretion increases in direct proportion to the concentration

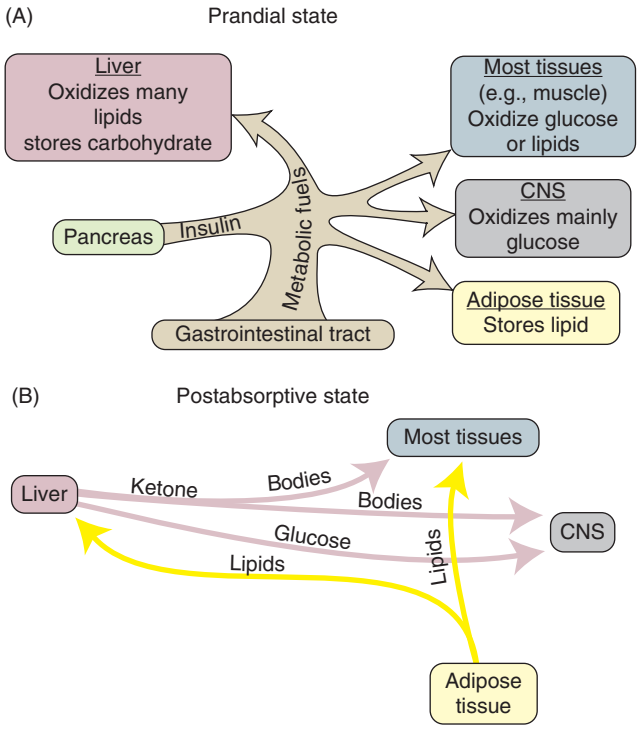


FIGURE 36.1 Schematic diagram of the fluxes of metabolic fuels in the (A) prandial and (B) postabsorptive states. Note the absence of insulin in the postabsorptive state, which greatly facilitates the mobilization of energy stores from the liver and adipose tissue. The adult human liver stores sufficient fuel to support metabolism during fasting for about 7 h, whereas the adipose mass has a far greater capacity.

of glucose in the blood. Other substrates, such as amino acids and ketone bodies, also stimulate insulin secretion. In addition, autonomic nerves innervate the pancreatic islets: cholinergic parasympathetic activity stimulates secretion of insulin, and α-adrenergic sympathetic activity inhibits it.

When a hungry person anticipates a meal, the aroma and, subsequently, the taste of food initiate the secretion of insulin and many other pancreatic and gastrointestinal signals via neural activity. Neural signals descend from the forebrain (where the smell and taste of food are recognized) through the hypothalamus to the dorsal motor nucleus of the vagus in the caudal brainstem and then to the pancreas and gastrointestinal tract by way of cholinergic fibers of the vagus nerve. This *cephalic phase* of secretion of insulin and gastrointestinal hormones such as glucagon-like peptide-1 (GLP-1) helps reverse the mobilization of fuels that occurs during fasting and prepares the body for the entry of fuels from the gut. As ingested food enters the stomach and duodenum, several gastrointestinal hormones, which also are important in digestion, continue to stimulate pancreatic B cells to secrete more insulin. This *gastrointestinal phase* of insulin secretion ensures that the level of insulin in the circulation is high by the time digested nutrients first appear

in the bloodstream. Finally, nutrients absorbed from the small intestine cause even greater stimulation of insulin secretion by their direct effect on the pancreas. This *substrate phase* of insulin secretion increases insulin levels further, and that effect lasts well beyond the cessation of eating. As a result of these coordinated meal-related events, prandial insulin secretion is rapid and appropriate to the caloric load, and ingested fuels are efficiently used and stored (Fig. 36.2).

Neural and endocrine factors control the ebb and flow of metabolic fuels from body stores. Circulating insulin is the most important factor that promotes storage. Insulin enables most tissues to take up glucose from the blood for immediate oxidation or for storage during the prandial period. Conversely, the most important factor for fuel mobilization is the near disappearance of insulin from the circulation during the postprandial, postabsorptive period, when fuel delivery from the small intestine has ended and parasympathetic activity to the pancreas is no longer elevated. Insulin secretion at this time is reduced greatly but is not inhibited and therefore can be stimulated immediately if necessary. For example, when more than the needed amount of stored fuels is mobilized, substrates directly stimulate the pancreas to secrete insulin, thereby slowing substrate mobilization to a rate appropriate to need. Thus, insulin is pivotal for storing calories in the fed state and for allowing stored calories to be mobilized in measured amounts during the postabsorptive state.

The amount of body fat (*adiposity*) is superimposed on other factors that influence insulin secretion. People with low adiposity have a relatively large number of active insulin receptors on adipose tissue, whereas obese individuals have fewer active insulin receptors. Consequently, in response to a given stimulus, secretion of insulin is lower in people who are lean than in those who are obese, but its net effects on adipose and other tissues are comparable. In healthy, nondiabetic people, this reciprocal relationship between the secretion of insulin and the sensitivity of most tissues to insulin ensures efficient storage and use of fuels independent of body fat. In these individuals, plasma insulin levels in the prandial and the postprandial periods are reliable correlates of adiposity.

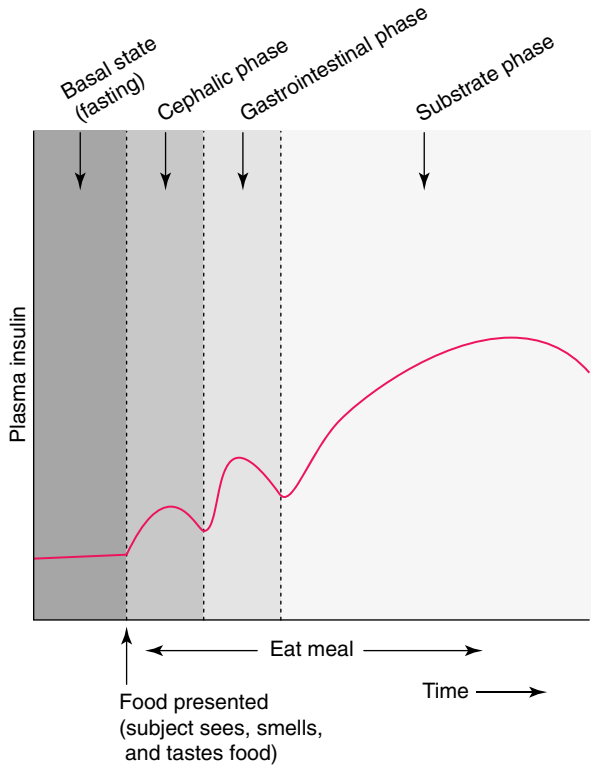


FIGURE 36.2 The three phases of meal-related insulin secretion. The cephalic phase, during which parasympathetic vagal activity stimulates the pancreas, is initiated when food is seen, smelled, and tasted. The gastrointestinal phase is mediated by the direct action of digestive hormones on the insulin-secreting B cells. The prolonged substrate phase is caused by metabolic fuels (mainly glucose) directly stimulating pancreatic B cells. When meals are prolonged, the three stimuli operate simultaneously and have additive effects.

Summary

All cells require continuous supplies of metabolic fuels to support ongoing activity. The fuels enter the circulation from the small intestine during the prandial period and from storage depots during fasting. The availability of fuels to tissues is controlled primarily by the liver and by the hormone insulin. Hepatic function and insulin secretion are in turn controlled in large part by the autonomic nervous system.

ROLE OF CALORIC HOMEOSTASIS IN CONTROL OF FOOD INTAKE

Meals Generate Biological Satiation Signals

The traditional view of caloric homeostasis is that animals eat when they need calories. However, as discussed earlier, the delivery of metabolic fuels to cells is continuous, and animals may experience urgent needs for calories to support cellular metabolism only infrequently. Instead, caloric homeostasis can be related to the control of food intake in ways that are unrelated to acute cellular needs.

Animals consume food in distinct bouts (i.e., meals); therefore, daily food intake reflects the cumulative intake of multiple meals. Control factors influence the time when each meal is initiated and the amount of food consumed before the meal is terminated. Satiation signals terminate an ongoing meal whereas satiety signals prolong the interval until a subsequent meal is initiated.

The first comprehensive study of meals was reported in 1966 by Le Magnen and Tallon. They maintained laboratory rats in cages and allowed them to eat *ad libitum* for weeks, during which time intake was recorded by photosensors and electronic relays. They consequently obtained a full record of when meals were initiated, how much was eaten during each meal, how much time passed before the next meal, and so on. No relationship was found between how much a rat ate in a meal and how much time had passed since it had last eaten; in other words, meal sizes were unpredictable. However, the larger the meal, the longer the interval before the next meal, and the relationship was strongest during the night, when rats ate the largest meals. Thus, eating appeared to be inhibited by a satiety signal generated in proportion to the size of a meal, and eating resumed when that signal disappeared. Similar observations have since been made in many laboratories.

Gastric Distension, Cholecystokinin, and Other Satiation Signals

Several meal-related variables might plausibly provide satiation signals. These include signals based on the smell, taste, and texture of food. Signals arising from the stomach could also play a role, as could intestinal and postabsorptive signals that arise after ingested food has left the stomach. In studies of rats, pregastric signals have been ruled out because little satiation is achieved when ingested food is immediately drained out through an esophageal or gastric fistula; these animals eat continuously, as if they had no satiation despite the passage of very large amounts of food through the oropharynx (Fig. 36.3). Because meals end long before significant digestion and absorption occur, gastric signals have been considered a likely source of satiation. Gastric volume is an obvious possibility, and, in fact, meals are known to end when substantial gastric distension occurs. The stomach wall is richly endowed with stretch receptors whose activity increases in proportion to the volume of the stomach. Those signals are communicated by way of the vagus nerve to the nucleus of the solitary tract (NST) and adjacent area postrema in the brainstem. The signals travel to the hypothalamus and, ultimately, to the cortex, where gastric distension is perceived.

Gastric stretch that accompanies meals presumably interacts with other signals to produce satiation. For example, aside from directly affecting digestion, the intestinal peptide *cholecystokinin* (CCK), which is secreted during meals, acts on receptors located on vagal afferent fibers that carry gastric stretch signals from the pyloric region of the stomach to the brainstem. Thus, relatively small amounts of CCK can inhibit feeding in rats by acting synergistically with gastric distension. When given in larger doses, CCK can inhibit food intake even when the stomach is empty

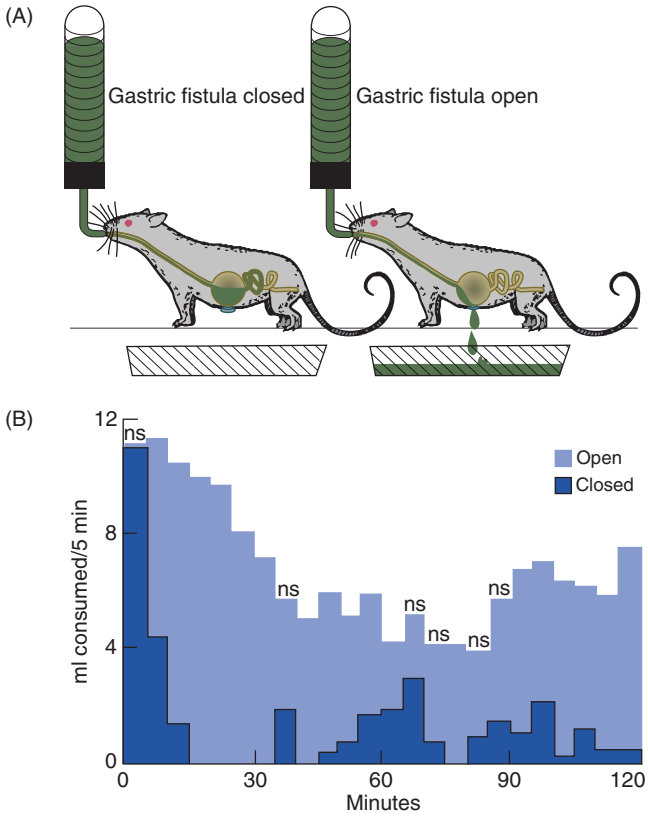


FIGURE 36.3 (A) Rats consuming liquid diet, with gastric fistula closed (real eating) or open ("sham eating"). When the gastric fistula is closed, food passes through the stomach and into the small intestine normally. When the gastric fistula is open, ingested food drains out the open fistula instead of accumulating in the stomach to distend it, and no food enters the small intestine. (B) Mean consumption of liquid diet (ml/5 min) by rats with a gastric fistula after 17 h of food deprivation. The rats consumed food in discrete meals when the fistula was closed. When the fistula was open, they ate continuously throughout the test period and never displayed satiety. Thus, rats consumed significantly more food during the 2-h test period when their fistulas were open than when they were closed. ns, intervals in which the amount of food consumed did not differ significantly between open and closed fistulas. *From Smith, Gibbs, and Young (1974).*

(Gibbs, Young, & Smith, 1973). As might be expected, total or selective gastric *vagotomy* (or ablation of brainstem areas to which the gastric vagus projects) decreases the ability of gastric volume and peripherally administered CCK to reduce meal size. Conversely, systemic administration of CCK receptor antagonists increases the size of meals (Box 36.1).

Gastric distension is one of the factors that control suckling in neonatal rats. The dam controls the timing of meals, and at each meal pups consume as much milk as their stomachs allow; gastric volume, not caloric content, provides the signal to stop. Consistent with the importance of gastric distension to suckling, pups are particularly sensitive to the inhibitory effects of

BOX 36.1

VALIDATING SATIATION FACTORS

Several metabolically important peptides are secreted during meals, including CCK, insulin, leptin, *glucagon-like peptide-1* (GLP-1), *Peptide YY* (PYY) and others. The specific individual nutrients being consumed are thought to determine the specific mix of peptides secreted, and their combination is considered to elicit satiation. Consistent with this perspective, each of these peptides has been found to reduce food intake when administered systemically to laboratory animals. By themselves, these findings do not prove that the peptides normally function as endogenous satiation agents; blood levels of an administered peptide must be within the physiological range. In addition, the behavioral effects of the peptide must be specific to the inhibition of food intake and cannot merely reflect a secondary consequence of illness, behavioral depression, or motor incapacitation. Because investigators cannot be certain what animals sense, they must infer whether animals experience satiation in association with an observed reduction in food intake.

One common approach is to determine whether the peptide causes a *learned flavor aversion* in animals. People readily learn to avoid food or drink that, when ingested, produces nausea. Rats and other animals seem to respond in the same way; moreover, when a toxin that elicits nausea is administered systemically soon after consumption of an uncontaminated drink with a novel flavor, the animals subsequently avoid fluids of that flavor and behave as if the drink had contained the toxic agent that had made them sick. In addition, electrophysiological recording from the first gustatory relay nucleus in the brainstem, the nucleus of the solitary tract, shows that the response elicited by a taste that has been associated with a toxin is similar to the pattern of activity typical of a naturally aversive flavor. Nausea is critical to the process; damage to the "emetic center" in the area postrema eliminates the sensation of nausea and prevents the formation of the learned taste aversion.

Another approach to distinguishing nausea from satiation is to monitor biological variables that occur in association with nausea but not satiation, or vice versa. One such variable is neurohypophyseal hormone secretion. For example, administration of lithium chloride (LiCl) causes nausea in humans and vomiting in monkeys, and elicits a learned flavor aversion in rats. LiCl stimulates vasopressin secretion in humans and monkeys. In this example, elevated plasma vasopressin can be considered a biological marker of nausea because it does not itself cause nausea or vomiting. Curiously, rats secrete the other neurohypophyseal peptide, oxytocin (and not vasopressin), in response to large doses of LiCl and other nauseants. Control experiments have shown that following an ordinary meal, vasopressin is not secreted in primates nor is oxytocin secreted in rats. Thus, when rats are administered a chemical agent that stimulates pituitary oxytocin secretion, nausea, rather than satiation, is the suspected basis of the reduction in food intake.

Such experiments are helpful for interpreting the anorexia that occurs when a peptide is administered exogenously, but insight into the normal effects of endogenously secreted chemical signals requires an alternative experimental approach. For example, a drug that blocks the effect of an endogenous peptide on its receptors could be administered. If the hormone normally functions to reduce food intake, then the blocking agent should increase the size of meals. Such drugs are not yet available for all peptides suspected of mediating endogenous satiation, but increases in meal size were found in experiments using drugs to reduce the activity of CCK, GLP-1, and several other peptides at their respective receptors. These observations strongly suggest that these peptides normally function as endogenous satiation factors.

CCK on food intake, suggesting that a combination of gastric distension and CCK may be the principal signals that inhibit the initiation of spontaneous meals in suckling rats.

As rat pups mature, signals related to the caloric content of food begin to participate in the control of meal size and become integrated with CCK, GLP-1, gastric distension, and other factors as satiation signals. This developmental change allows adults to ingest, assess, and respond to foods whose caloric density is not as constant as that of milk. For example, when liquid food is diluted with water, adult rats compensate by increasing the volume consumed in each meal, allowing daily caloric intake to remain stable. This observation reveals that gastric distension is but one of several possible satiation signals in adult rats and that greater distension can be accommodated in circumstances in which the calorie-related satiation signals have diminished.

The mechanism by which caloric signals are monitored is an area of intense investigation. Although early reports suggested that the stomach monitored caloric nutrients, later findings did not support this hypothesis. For example, when hungry rats were equipped with closed pyloric cuffs to prevent ingested food from entering their small intestine, the rats decreased their food intake in proportion to the volume, and not the caloric

content, of intragastric infusions (Phillips & Powley, 1996). In contrast, when the cuff was open so that gastric contents could empty into the duodenum, caloric nutrients were much more effective in reducing food intake than calorie-free loads of equal volume. These observations suggest that the stomach does not detect caloric nutrients but instead contributes primarily inhibitory signals related to gastric distension.

Postgastric Effects of Meals Provide Additional Satiation Signals

During the course of a meal, some of the ingested food enters the small intestine and is absorbed, thus allowing postgastric signals to contribute to satiation. In this regard, small amounts of specific nutrients infused directly into the duodenum produce a robust suppression of eating that cannot be attributed to gastric factors. It is likely that both the upper small intestine as well as postabsorptive sites are responsible for detecting the administered calories and generating the signals that limit meal size. The luminal surface of the intestine contains mainly enterocytes that absorb digested nutrients and pass them into the blood. Interspersed among the enterocytes are scattered enteroendocrine cells that have tentacle-like projections which extend into the intestinal lumen and express many of the same taste receptors as occur on taste buds on the tongue. As a result, these enteroendocrine cells "taste" specific nutrients (e.g., glucose) that are in the intestinal lumen, and they respond by secreting peptide hormones that contribute to the ongoing process of digestion. There are a dozen or so different types of enteroendocrine cells, each secreting a different peptide or peptides. As an example, I-cells secrete CCK when they "taste" fatty acids in the lumen, and the CCK in turn enters the blood and causes the liver/gall bladder to secrete bile into the intestine to facilitate the digestion of fats. Analogously, L-cells secrete GLP-1 and PYY in response to "tasting" carbohydrates, and GLP-1 increases insulin secretion during the meal. Importantly, many of the peptide secretions of the enteroendocrine cells also double as satiation signals, acting either locally on afferent nerves in the wall of the intestine or else circulating to a distant organ such as the liver or brain to elicit satiation. These peptides include CCK, GLP-1, and PYY.

As discussed earlier, the liver plays a critical role in caloric homeostasis, and the delivery of nutrients absorbed from the small intestine to the liver may be monitored to provide an important postgastric signal for the inhibition of food intake. Consistent with this hypothesis are findings that infusion of glucose or lipids into the hepatic portal vein reduces food intake. Importantly, infusion of fructose (a sugar utilized by the liver but not by the brain) into the general circulation no longer reduces food intake in rats after hepatic vagotomy. Collectively, these findings suggest that like the intestines themselves, the liver provides a satiation signal. All of these signals would diminish as absorption slows and satiation occurs.

Intravenously administered glucose reduces food intake but not by as much as the caloric content of the glucose. In contrast, when glucose solution is consumed orally or fed through a tube directly into the stomach, the compensatory reduction in food intake is equivalent to the caloric load. In this case, gastric and/or intestinal signals such as GLP-1 contribute to the increased satiation and augment insulin secretion above that elicited by the glucose *per se*; that is, more insulin is secreted in response to infusion of glucose into the stomach or intestine than into a vein, which might contribute to the observed reduction in appetite. In fact, intravenously infused glucose reduces food intake by an equivalent caloric amount when insulin is added to the infusate. Together, these observations suggest that insulin contributes to the satiating effect of glucose whether given by gavage or ingested during a meal. Consistent with this possibility, rats made diabetic by destruction of their pancreatic B cells eat discrete meals more frequently as if they experienced less postprandial satiety (Box 36.2).

In addition to providing useful nutrients, consumption of food increases *plasma osmolality*, yet another factor influencing the size of a meal. Increased plasma osmolality stimulates thirst, and for this reason water intake usually accompanies eating. The important point, however, is that less food is consumed during meals when drinking water is not available, a phenomenon known as *dehydration anorexia*. Similarly, increased plasma osmolality caused by systemic injection of hypertonic saline reduces food intake in proportion to the administered osmotic load. The well-described forebrain osmoreceptors that cause thirst and influence vasopressin secretion (Chapter 39) apparently do not mediate this inhibitory effect on food intake, which persists after those osmoreceptors have been destroyed surgically. Instead, osmoreceptors in the viscera or in the caudal brainstem might mediate the inhibitory effect of increased plasma osmolality on food intake.

Body Weight Influences Food Intake

The control of food intake is also associated with the maintenance of body weight (adiposity, actually). For example, after a period of food deprivation and forced loss of body weight, animals (including humans) eat relatively large meals until adiposity returns to pretreatment levels. Conversely, after a period of forced feeding, during which the increased daily intake of calories causes weight gain, animals eat relatively small meals

BOX 36.2

THE GLUCOSTATIC HYPOTHESIS

Almost 60 years ago, Mayer (1955) proposed that the onset and termination of meals are determined by changes in the amounts of glucose used by certain areas of the brain. More specifically, Mayer hypothesized that food intake was controlled by a region of the ventral hypothalamus whose glucose utilization, unlike the rest of the brain, was sensitive to insulin. In diabetic rats, the presence of hyperphagia despite elevated blood glucose emphasized the point that diminished glucose utilization in these insulin-dependent cells (as opposed to a reduction of blood glucose levels *per se*) was the critical factor in removing satiety and stimulating food intake.

At first widely accepted, the glucostatic theory fell into disfavor for several reasons. For one, although eating can be induced by low availability of glucose to the brain, no evidence supports suppression of eating by high glucose availability. In addition, concentrations of blood glucose after administration of large doses of insulin are now known to be far lower than what occurs during a fast or at the normal onset of a meal. Thus, eating in response to insulin-induced hypoglycemia is now considered an "emergency response" that is not relevant to normal feeding. (A similar emergency response occurs after a systemic injection of 2-deoxyglucose induces an acute decrease in cellular glycolysis.) Finally, the prevailing view of the ventromedial hypothalamus as a satiety center has changed considerably since the original formulation of the glucostatic hypothesis. This region of the brain is better regarded as a regulatory control center that influences many aspects of hormonal and autonomic function. Diabetic hyperphagia may therefore represent the ineffectiveness of insulin or leptin to activate neural systems that inhibit food intake. In contrast, the earlier glucostatic hypothesis would have explained diabetic hyperphagia as a consequence of the decreased uptake of glucose into special hypothalamic neurons that mediate satiety.

In recent years, evidence has emerged that neurons in the hypothalamic arcuate nucleus are sensitive to fluctuations of the adiposity signals leptin and insulin. Neurons in the same area also are sensitive to fluctuations of glucose as well as to certain fatty acids (especially oleic acid) and amino acids (especially leucine), and local manipulations of either adiposity signals or nutrients influence food intake, pancreatic insulin secretion, and the release of glucose from the liver. Hence, the view is emerging that the arcuate nucleus integrates long-term energy stores (principally body fat) with acute availability of nutrients to control several aspects of energy homeostasis.

Stephen C. Woods and Edward M. Stricker

Reference

Mayer, J. (1955). Regulation of energy intake and body weight: The glucostatic theory and the lipostatic hypothesis. *Annals of the New York Academy of Science, 411,* 221–235.

(or no meals at all) until normal adiposity is restored. Long-term stability of adiposity in adult animals is attributed in part to these compensatory responses to weight fluctuations that are caused by acute changes in food intake or energy expenditure.

Adiposity may indirectly influence food intake by modulating how quickly food passes through the gastrointestinal tract, how long nutrients remain in the circulation, or how nutrients interact with the liver. For example, after fasting, less prolonged gastric distension or more rapid absorption and storage of ingested calories would diminish the duration of satiety signals, thereby increasing the frequency of meals. These effects would promote the *hyperphagia* (overeating) associated with the restoration of body weight in animals that have been food restrained and lost weight. Similarly, the satiating properties of ingested food would be reduced by diversions of ingested food calories to the fetus during pregnancy, to milk during lactation, or to skeletal muscle during exposure to cold temperatures, thereby contributing to the hyperphagia associated with each of these conditions.

Alternatively, adipose tissue may directly signal the brain to modulate food intake. Although afferent nerves from adipose tissue to brain have been described, communication from adipose tissue is thought to be mediated primarily by a humoral (circulating) factor. *Parabionts*, created by surgically joining two animals at the flank muscles and skin, support this hypothesis. Vascular interconnection within a parabiont is demonstrated when a dye injected into one animal appears in the blood of its partner. Because the nervous systems of the two animals remain independent, any physiologic communication between the animals must occur through the vascular system. When two lean rats were joined parabiotically, each consumed its normal amount of food and maintained its usual body weight. Similar results were obtained when two obese rats with lesions in the *ventromedial hypothalamus* (see later) were joined together. However, when a lean animal was joined to an obese

animal, the result was striking: The lean animal reduced its food intake markedly and consequently lost a considerable amount of weight. These experiments have been interpreted to mean that a circulating factor, secreted by the obese animal in large amounts proportional to adiposity, enters the blood of the lean animal and reduces food intake, as if the lean animal was receiving a signal that it was too fat (Coleman & Hummel, 1969).

Investigations have identified *leptin*, a hormone synthesized by adipose tissue, as a key circulating adiposity signal that regulates food intake and body weight. Leptin concentration in plasma is directly proportional to adiposity, and a receptor-mediated transport process passes it into the brain. Animals that lack the gene for synthesizing leptin (ob/ob mice) are hyperphagic and obese, and administering leptin to ob/ob mice causes them to eat less and lose weight. Because lower doses of leptin have this action when administered into the ventricles of the brain than when administered systemically, leptin is believed to reduce food intake and body weight via a central site of action. Consistent with this hypothesis, the receptor for leptin has been identified and found to be located within the hypothalamus. Like ob/ob mice, animals with mutated leptin receptors (db/db mice and fatty Zucker rats) are obese, but unlike ob/ob mice, they do not respond to exogenous leptin. Hence, leptin appears to function as a negative feedback signal to the brain. When fat stores increase in adipose tissue, more leptin is secreted and enters the brain, causing greater inhibition of food intake and loss of body fat. When circulating leptin levels are low, feeding and other anabolic responses are disinhibited. In this way, leptin acts to promote the maintenance of relatively stable adiposity over long intervals. The neural circuitry through which leptin inhibits food intake and influences caloric homeostasis is discussed later.

Insulin provides a second circulating signal that informs the brain of body fat levels. Plasma levels of insulin are directly correlated with adiposity, and, like leptin, insulin is transported through the blood–brain barrier where it can act on insulin receptors in the hypothalamus and other sites. Insulin administered into the brain reduces food intake and body weight, whereas animals that do not secrete insulin are hyperphagic. Of course, severe diabetics do not become obese because their adipocytes cannot store fat in the absence of insulin. Mice lacking insulin receptors on brain cells are both hyperphagic and obese. In short, both leptin and insulin provide blood-borne signals that enter the brain and act on their respective receptors to reduce food intake and body fat.

Obesity, the accumulation of greater than normal fat in the body, is increasing throughout the world and creating a major public health problem, in part because it increases the risk for developing cardiovascular disorders, several cancers, type-2 diabetes mellitus, and other metabolic problems, and increases mortality. Obese individuals have both leptin resistance and insulin resistance; that is, although they secrete large amounts of both leptin and insulin in proportion to their adiposity, these signals are relatively ineffective in controlling metabolic parameters. Consequently, considerably more insulin is required to maintain plasma glucose within normal limits, and the brain is insensitive to the catabolic action of both leptin and insulin. The latter is thought to be due to a combination of reduced passage of these hormones through the blood–brain barrier as well as fewer leptin and insulin receptors within the brain itself. See Box 36.3 for more on obesity.

BOX 36.3

TREATMENT OF OBESITY

Obesity, the accumulation of greater than normal fat in the body, results when energy intake exceeds energy expenditure over long intervals. Obesity often is defined in terms of the *body mass index* (BMI), which is weight in kilograms divided by height in meters squared. A BMI of 18.5 to 24.9 is considered normal, with <18.5 being underweight, 25 to 29.9 being overweight, and >30 being obese. A BMI over 40 is morbid obesity. However, the calculation of BMI does not consider the distribution of fat within the body, yet this is recognized as a major determinant of health risk. Abdominal fat (mainly around the viscera) is more metabolically active, more highly vascularized and innervated, and more likely to be a site of inflammation in the body than subcutaneous fat; abdominal fat (typified by the "potbelly" or "apple" stereotype) therefore carries a much higher health risk than subcutaneous fat (characterized as a "pear" shape). Because of this, assessment of the circumference of the waist (abdomen) at the top of the iliac crest is thought to be a better predictor of health risk than the BMI. By this determination, obesity is defined as a waist circumference of >102 cm in men and >88 cm in women.

The incidence of obesity is increasing throughout the world and creating a major public health problem, in part because it increases the risk for developing cardiovascular disorders, several cancers, type-2 diabetes mellitus, and other metabolic problems, and increases mortality. Obese individuals have both leptin resistance and insulin resistance; that

BOX 36.3 (*cont'd*)

is, although they secrete large amounts of both leptin and insulin in proportion to their adiposity, those signals are relatively ineffective at controlling metabolic parameters. Consequently, considerably more insulin is required to maintain plasma glucose within normal limits, and the brain is insensitive to the catabolic action of both leptin and insulin. The latter is thought to be due to a combination of reduced passage of these hormones through the blood–brain barrier as well as fewer leptin and insulin receptors within the brain itself.

Hypotheses as to why the incidence of obesity has been increasing so substantially in recent years abound, and often presume some combination of increased availability of energy-dense palatable foods and/or decreased average energy expenditure. It is axiomatic to state that obesity can be prevented or cured by reversing the trend—that is, by eating less and exercising more. Unfortunately, many humans in modern society cannot realistically modify these behaviors for long intervals, and they have consequently turned to other methods to aid in weight loss. Historically, several different neuropharmacologic approaches have been used, most of them targeting brain neurotransmitter systems thought to be important in appetite control. Specifically, pharmaceuticals developed to treat obesity have targeted the serotonin, catecholamine, or endocannabinoid systems, either with ligands for their respective receptors or else with chemical agents that influence their secretion, degradation, or reuptake. However, because each of these transmitter systems influences multiple brain circuits, and because none of the compounds available to date have been appetite-specific and often have undesirable side effects on mood and other behaviors, none are currently approved by the FDA in the United States to treat obesity. Further, actual weight loss achieved any by these pharmacological approaches was rarely greater than 10%.

During the last decade, bariatric surgery has become the treatment of choice, especially for morbidly obese individuals, and these surgeries often result in far greater weight loss than that achieved pharmacologically (i.e., up to 70% of excess fat or more). In the conceptually simplest procedure, a band is placed around the stomach near the esophagus. The result of this gastric banding procedure is that food is swallowed into a very small gastric pouch, severely limiting the amount of food that can be consumed at once due to gastric restriction. Distension of the remaining gastric pouch is thought to result in smaller meals. Even greater loss of weight may occur with more invasive procedures that involve rearrangement of the normal routing of ingested food through the gastrointestinal system. In a roux-en-Y gastric bypass (RYGB), swallowed food enters a small gastric segment and then is immediately shunted into the distal small intestine, effectively bypassing half or more of the small intestine and theoretically reducing the absorptive area for food. In a vertical sleeve gastrectomy (VSG), most of the stomach is surgically removed and consequently swallowed food enters a small cylinder of remaining stomach leading directly into the small intestine, thereby eliminating most of the stomach's normal ability to store ingested food. These commonly performed procedures result in rapid, substantial, and comparable weight loss; for example, many patients have their diabetes greatly improved, often within days after the surgery.

The mechanisms for the weight loss and reduction of diabetic symptoms after RYGB or VSG are not well understood. Although increased gastric stretch may be a factor in some procedures, it is not common to all, and malabsorption seems to have a minimal effect at best. Instead, a major contributing factor may be an alteration of the normal neural and endocrine signaling between the GI tract and the brain. For example, one feature common to the successful procedures is greatly increased meal-induced secretion of peptides such as GLP-1 and PYY from enteroendocrine cells, perhaps thereby creating a greater satiation signal as well as a greater stimulation of insulin secretion during meals. In this regard, many diabetic patients lose substantial amounts of body weight when their pharmacotherapy includes GLP-1 agonists.

Summary

Three effects of eating limit the size of an ongoing meal: gastric distension (potentiated by CCK), postgastric detection of calories (via satiation signals such as CCK and GLP-1, perhaps potentiated in the liver by insulin), and increased plasma osmolality. These signals reach the brain through visceral afferent fibers (especially those traveling in the vagus nerve) and the circulatory system. Meal size invariably increases after the inhibitory effects are experimentally removed by blocking the detection of gastric stretch or of satiating peptides, diluting the food with noncaloric material, or rehydrating the animal. Inhibition appears to be integrated so that as one effect diminishes (e.g., the signal associated with gastric distension), another increases (e.g., the signal associated with the postabsorptive delivery of calories and insulin to the liver), thereby maintaining satiety and prolonging the interval between meals. When the satiety signals disappear, hunger emerges and stimulates initiation of another meal.

Food intake is also normally linked closely with body weight. In experimental animals, periods of starvation or forced feeding are followed by self-regulated compensatory changes in eating patterns until prior body size is regained. Links between body weight and food intake can be mediated indirectly by altered gastrointestinal motility as well as by the blood-borne, centrally active hormones leptin and insulin.

CENTRAL CONTROL OF FOOD INTAKE

Early theories of the control of food intake focused on signals derived from stomach and blood. Peripheral factors, such as the concentration of glucose in the blood, were plausibly associated with food intake and were accessible to experimentation. Theories involving the brain emerged after development of a stereotaxic instrument that enabled discrete lesions of the cerebrum to be made in laboratory animals, permitting evaluation of ideas about brain function (Fig. 36.4). The first findings from such experiments were striking: Bilateral electrolytic lesions of the ventromedial hypothalamus (VMH) caused marked hyperphagia and obesity in rats. This observation was interpreted to mean that the animals had become less sensitive to incoming signals of satiation and that they therefore overate and became fat. Bilateral lesions in the adjacent *ventrolateral hypothalamus* (VLH) were later found to cause *aphagia* (absence of eating), which led to death by starvation. This finding was interpreted to mean that these rats no longer detected hunger signals, and thus they starved to death while unaware of their internal state. Collectively, these results provided the foundation of a dual-center hypothesis, in which a satiety center in the VMH was thought to suppress activity in a hunger center in the VLH. Both the syndrome of hyperphagia and obesity and the syndrome of aphagia and weight loss that result from focal hypothalamic lesions have been reproduced in multiple laboratories and in multiple species. However, additional observations have caused reinterpretation of these familiar and reliable findings, and they now have a very different meaning.

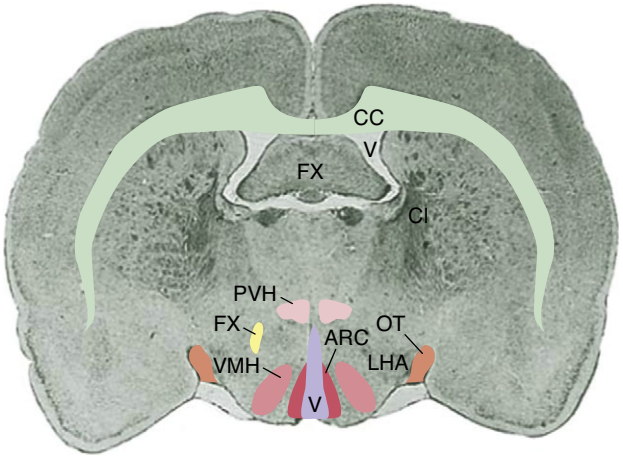


FIGURE 36.4 Coronal section through an adult rat brain at the level where the optic tract enters the two hemispheres. The lateral ventricle and third cerebral ventricle (V) are shaded, and selected fiber tracts are outlined for reference (OT, optic tract; CC, corpus callosum; FX, fornix). Key hypothalamic sites important in caloric homeostasis, present in each hemisphere, are paraventricular nuclei (PVN), ventromedial nuclei (VMN), arcuate nuclei (ARC), and the lateral hypothalamic area (LHA).

Food Intake is Not Controlled by Hypothalamic Hunger and Satiety Centers

The strength of the traditional dual-center model, in which one center mediates hunger and the other satiety, was that it provided a simple answer to the question of how the brain controls intake of food. The principal limitation of the model, however, was that it could not answer the next level of questions: What signals control individual meals? How are these signals integrated with physiological aspects of caloric homeostasis, including long-term maintenance of body weight? What is the influence of other neural sites and systems in the brain on the control of food intake? How are excitatory and inhibitory influences on eating integrated? Investigators are addressing these and related questions, but they now usually take a perspective that does not presume the existence of brain centers mediating hunger and satiety.

In addition to causing hyperphagia, VMH lesions profoundly reduce sympathetic tone and increase parasympathetic tone and vagal reflexes. For example, the high concentration of insulin in animals with VMH lesions is not simply a secondary consequence of hyperphagia but occurs even when food intake is limited. Because of the change in autonomic tone, caloric equilibrium in adipose tissue shifts away from lipolysis and toward lipogenesis, thereby allowing more rapid storage of ingested calories after a meal. Increased parasympathetic activity also allows faster gastric emptying. Importantly, VMH lesions do not cause loss of satiety. Meal sizes continue to correlate well with postmeal intervals, but the interval after a meal of any size is shorter in animals with VMH lesions than in control animals. Because the duration of their postprandial satiation signals is shorter than normal, rats with VMH lesions eat more frequent meals. They remain hyperphagic until the equilibrium between lipogenesis and lipolysis is reestablished in adipocytes. For this to happen, triglyceride stores must accumulate in the cells to levels sufficient to create insulin resistance, whereupon lipogenesis subsides, gastrointestinal motility diminishes, meals are eaten less frequently, and daily food intake is

reduced. The animal maintains its obese state as if it had a new "set point" for body weight. Thus, after obese rats with VMH lesions have been deprived of food and forced to lose weight, hyperphagia reoccurs until the prior level of high adiposity is reattained. Similarly, when neurologically normal rats are overfed and become obese prior to receiving a lesion of the VMH, they are not hyperphagic after the lesion but rather eat to maintain their elevated weight.

In sum, rats gain body fat after VMH lesions, and as they are becoming obese (indeed, in part because they are becoming obese), the duration of satiety after meals is relatively low. However, as fat accumulates the duration of satiety after meals increases back toward normal. This explanation contrasts with early hypotheses that body weight increased after VMH lesions because animals became obligatorily hyperphagic. The primary phenomenon is now recognized as a change in autonomic tone to promote vagal reflexes and fat storage. Because that change is observed even when increased food intake is prevented, hyperphagia cannot be the primary phenomenon. The key element appears to be the chronic, neurally stimulated increase in insulin secretion. When the increased insulin secretion is prevented by denervating the pancreas, obesity does not develop and hyperphagia is greatly attenuated.

In contrast to VMH lesions, rats with large VLH lesions do not eat and may starve to death. However, they also do not drink water when dehydrated and they do not move about in their cages or respond to diverse stimulation. In other words, the most prominent abnormalities in these rats are akinesia and sensory neglect. In this respect, rats with VLH lesions resemble human patients with Parkinson's disease, a neurological disorder that has been attributed to the degeneration of dopamine-containing neurons of the nigrostriatal bundle (see Chapters 30 and 31). Dopaminergic fibers course through the internal capsule just lateral to the VLH as they ascend from the ventral mesencephalon to the striatum along the medial forebrain bundle (Chapter 43). Large electrolytic lesions of the VLH area interrupt these dopaminergic fibers. More selective damage to the dopaminergic neurons near the VLH by intracerebral administration of the neurotoxin 6-hydroxydopamine also produces akinesia and sensory neglect in association with loss of food intake, and it does so without disturbing parasympathetic reflexes. Thus, the aphagia induced by large VLH lesions does not result from damage to a putative hunger center in the brain but instead reflects a more general disruption of movement and sensorimotor integration (Ungerstedt, 1971), at least in part (see later).

Summary

The ability of brain circuitry to control food intake initially was based on the consequences of stereotaxic lesions of ventromedial and ventrolateral hypothalamic nuclei leading to hyperphagia or aphagia, respectively. Subsequent studies revealed that each lesion also induces changes in autonomic function, in gastrointestinal activity, and in the equilibrium between lipolysis and lipogenesis in adipocytes. In addition, these lesions affected eating behavior through influences on spontaneous mobility and sensory responsivity to food-derived stimuli.

NEUROPEPTIDES AND THE CONTROL OF FOOD INTAKE

Although the details of the central control of food intake are incompletely understood, many neuropeptides have been implicated. To simplify the discussion, we focus on two broad categories of neuropeptides: those that are mainly anabolic and those that are mainly catabolic. Both of these subgroups in turn are modified by multiple influences, including satiation signals, adiposity signals, experience, habit, emotional state, the social situation, and so on. Anabolic peptides cause increased eating, decreased energy expenditure, and, when chronically elevated, increased body fat accumulation. Catabolic peptides reduce food intake, increase energy expenditure, and lead to loss of body fat. Figure 36.5 presents a simplified model of some of the better-known components of this control system. An interesting feature of the model is that major efferent outputs from the hypothalamus to autonomic nuclei and to motor nuclei controlling ingestion *per se* emanate from the VMH and VLH, the central neural sites that were the main focus of attention over a half-century ago.

The Hypothalamic Arcuate Nucleus (ARC) Mediates the Effect of Adiposity on Food Intake and Body Weight

The amount of body fat is signaled to the brain by leptin and insulin. Receptors for both peptides are located (among other sites) in the ARC on two distinct groups of neurons. The first group of ARC neurons synthesizes the neuropeptides, *α-melanocyte-stimulating hormone* (α-MSH) and *cocaine–amphetamine-related transcript* (CART); these neurons are activated by leptin and insulin. α-MSH and CART are potent catabolic peptides, and when either is administered locally into the third cerebral ventricle animals eat less food, have increased energy expenditure, and lose weight (Cone, 1999). α-MSH is in the *melanocortin* family of peptides, a group that also includes *adrenocorticotrophic hormone* (ACTH). These peptides act on melanocortin (MC) receptors, and two of these receptors, termed MC3 and MC4, are expressed in several nuclei of the hypothalamus, including the *paraventricular nucleus*

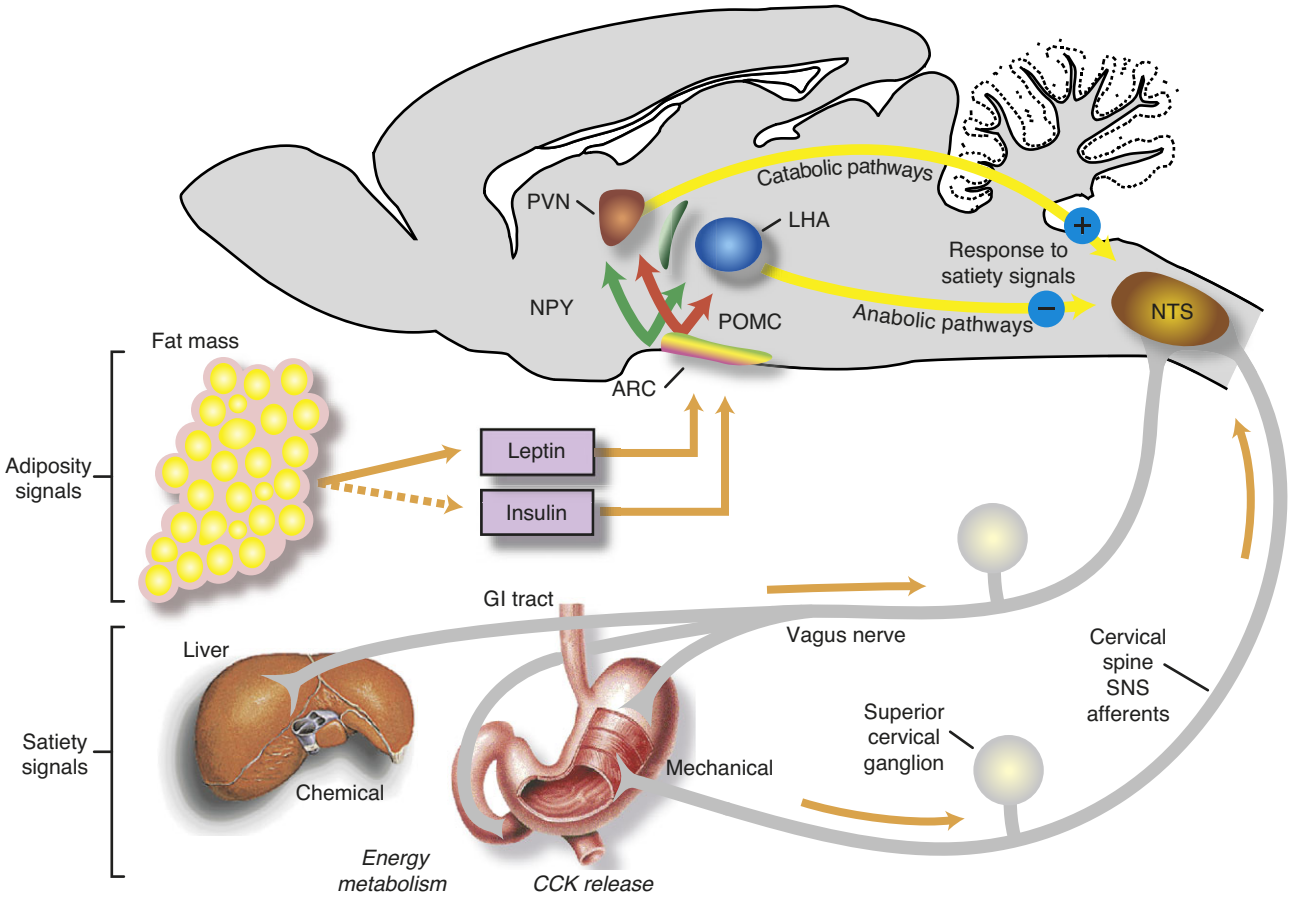


FIGURE 36.5 Schematic diagram of the signals that control caloric homeostasis. Satiety signals arising in the periphery such as gastric distension and CCK are relayed to the nucleus of the solitary tract (NTS) in the brainstem. Leptin and insulin, the two circulating adiposity signals, enter the brain and interact with receptors in the arcuate nucleus (ARC) and other brain areas. These adiposity signals inhibit ARC neurons that synthesize NPY and AgRP (NPY cells in the diagram) and stimulate neurons that synthesize proopiomelanocortin (POMC), the precursor of α-MSH, and CART. These ARC neurons in turn project to other hypothalamic areas, including paraventricular nuclei (PVN) and the lateral hypothalamic area (LHA). Catabolic signals from the PVN and anabolic signals from the LHA are thought to interact with the satiety signals in the brainstem to determine when meals will end. *From Schwartz, Woods, Porte, Seeley, and Baskin (2000).*

(PVN), VMH, and VLH, where α-MSH-containing axons project from the ARC and where α-MSH acts as an agonist. Animals that lack either MC3 or MC4 receptors become obese; animals with no MC3 receptors gain weight gradually over their lifetime without overeating, whereas animals without MC4 receptors are hyperphagic and become much fatter. The administration of selective MC4 antagonists causes neurologically intact animals to overeat and, if prolonged, to become obese. Hence, hypothalamic melanocortins, exemplified by α-MSH, exert a tonic catabolic effect to keep body weight from increasing (Cone, 1999). Importantly, the ability of exogenous leptin to reduce food intake and body weight is completely attenuated in the presence of an antagonist to MC3 and MC4 receptors, suggesting that the "downstream" pathway by which adiposity influences caloric homeostasis is via melanocortin signaling (Seeley et al., 1997).

The other type of ARC neuron influenced by adiposity signals synthesizes the neuropeptides *neuropeptide Y* (NPY) and *agouti-related protein* (AgRP). Both peptides are potent anabolic compounds in that the administration of either into the third ventricle results in hyperphagia, reduced energy expenditure, and weight gain. Although NPY is synthesized in many areas of the brain including the dorsomedial nucleus of the hypothalamus (DMN), AgRP synthesis is limited to the ARC. Neurons containing NPY and AgRP express receptors for both leptin and insulin, and the local administration of either insulin or leptin near the ARC reduces the synthesis of both NPY and AgRP. Hence, adiposity signals exert net catabolic effects through the joint mechanisms of stimulating α-MSH/CART neurons and simultaneously inhibiting NPY/AgRP neurons. For all these reasons, the ARC can be considered to be the brain's sense organ that detects body adiposity by monitoring

the levels of leptin and insulin. In turn, axons from these two groups of ARC neurons innervate many other hypothalamic nuclei as they modulate aspects of caloric homeostasis.

Two important tracts convey NPY-containing axons to the PVN, one from the ARC that co-expresses AgRP and one from the DMN that has no AgRP. The PVN expresses receptors for NPY, and careful mapping studies have identified the PVN as the most sensitive site to the orexigenic effect of exogenously administered NPY. Consistent with this finding, when antisense oligonucleotides to NPY receptors are given locally within the PVN, rats that have been deprived of food eat less than they usually would. Among the numerous sites that synthesize NPY within the brain, only those in the arcuate nucleus are sensitive to changes of food intake as signaled by leptin and insulin (Schwartz et al., 2000). NPY mRNA is increased in the ARC and DMN when animals are fasted and returns to baseline upon refeeding. Fasting also increases the levels of NPY in axon terminals in the PVN and increases secretion of NPY within the PVN. Hence, the fasted animal is primed to eat more food due in part to the action of elevated NPY in the area of the PVN. Circulating insulin and leptin are decreased during fasting, and local administration of either into the brain of a fasted rat lowers the elevation of NPY mRNA in the ARC and lowers the amount of food eaten by food-deprived rats (Schwartz et al., 2000). Hence, the metabolic state, as reflected by leptin and insulin, has a marked influence on arcuate NPY-containing neurons and consequently on neural pathways that influence energy intake and expenditure. Endogenous levels of NPY in the ARC–PVN system normally peak when daylight ends and nocturnal activity begins (Leibowitz, 1990), which also is the time when rats typically eat their largest meal of the day. Hence, NPY activity in the PVN appears to be a key mediator of anabolic activity.

The other neuropeptide synthesized in NPY neurons in the arcuate nucleus, AgRP, is an endogenous antagonist of MC3 and MC4 receptors (Cone, 1999). AgRP therefore promotes food intake by acting at MC receptors in the PVN and other areas to block the action of α-MSH. When exogenous AgRP is administered into the third cerebral ventricle, the resultant hyperphagia lasts for up to six days (Hagan et al., 2000). In contrast, the hyperphagia elicited by NPY lasts only a few hours. Hence, activation of the arcuate NPY/AgRP neurons results in an acute but robust increase of food intake caused by the actions of NPY and a steady and more prolonged increase caused by the actions of AgRP. Furthermore, the two peptides act in different ways. NPY acts through its receptors to stimulate anabolic pathways directly, whereas AgRP acts through a different receptor by antagonizing tonically active catabolic peptides. It is not known whether NPY and AgRP from ARC neurons act on the same or different target cells in the PVN, VLH, and elsewhere. An important distinction is that whereas NPY originating in both the ARC and the DMN project to the PVN and VLH, only DMN NPY neurons project to areas of the brainstem where satiation signals are relayed from the vagus (Moran, 2010).

In sum, the ARC appears to be a chemosensor in the brain that detects body adiposity by means of insulin and leptin receptors. One type of ARC neuron synthesizes peptide transmitters (α-MSH and CART) having a net catabolic effect, whereas another type synthesizes peptide transmitters (NPY and AgRP) with a net anabolic effect. Both types of neuron project to other hypothalamic areas where they influence specific aspects of caloric homeostasis.

Other Hypothalamic Peptides also Influence Food Intake

Several other neuropeptides important in the control of caloric homeostasis are synthesized in the VLH, including the *orexins* and *melanin-concentrating hormone* (MCH) (Shimada, Tritos, Lowell, Flier, & Maratos-Flier, 1998). Administration of either of these peptides into the brain stimulates food intake, whereas mice that lack the gene for MCH are lean and obesity resistant. Administering orexins into the brain also elicits wakefulness, and as orexin levels decline after a meal, wakefulness wanes and drowsiness takes over. It appears likely that the aphagia exhibited by animals with VLH lesions is due in part to reduced levels of these peptides and that the hyperphagia caused by electrical stimulation of the VLH is due in part to release of orexins and MCH.

Oxytocin is a peptide synthesized in the PVN and supraoptic nuclei of the hypothalamus and is secreted from the posterior lobe of the pituitary. In the systemic circulation of female mammals, this hormone is well known for stimulating uterine contractions for parturition and stimulating milk letdown in mammary glands during lactation. However, the hypothalamus and pituitary of males contain as much oxytocin as those of females; therefore, the peptide has long been suspected of being secreted in other circumstances and having additional functions. In fact, research has revealed that gastric distension, CCK administration, and plasma hyperosmolality, treatments that decrease food intake in rats, each elicit pituitary oxytocin secretion in male and female rats. In addition, increases in plasma concentrations of oxytocin correlate highly with observed decreases in food intake.

Although observations were consistent with oxytocin being an appetite suppressant, later experiments showed

that this was not the case: Systemic administration of oxytocin does not affect food intake. Instead, the reduced food intake seen after treatment with CCK or hypertonic saline appears to be mediated by oxytocin secreted from neurons whose cell bodies are located in the hypothalamic PVN but whose axons project within the central nervous system rather than to the pituitary gland. Consistent with this hypothesis, central injection of oxytocin into the cerebral ventricles (icv) decreases food intake in rats, an effect that is blocked by icv pretreatment of the animals with an oxytocin receptor antagonist. More to the point, the effects of exogenous CCK and hypertonic saline on food intake are blunted by icv pretreatment with an oxytocin receptor antagonist. A summary of some of the hypothalamic peptides that modulate food intake is depicted in Figure 36.6.

Stress is also well known to decrease food intake in animals, and various stressors activate the hypothalamic–pituitary–adrenal (HPA) axis. A key mediator of this effect is *corticotropin-releasing hormone* (CRH), which is synthesized and secreted from the PVN. CRH stimulates the secretion of ACTH from the anterior pituitary, and ACTH in turn elicits steroid hormone secretion from the adrenal cortex (see Chapter 40). In rats, icv injection of CRH decreases food intake in the absence of stress and in association with oxytocin secretion from the posterior pituitary. In addition, icv pretreatment with an oxytocin receptor antagonist eliminates CRH-induced inhibition of food intake. These findings are consistent with central oxytocinergic neurons mediating the inhibition of food intake that occurs during stress. Such inhibitory effects on eating would complement the known inhibitory effects of central oxytocin on gastric motility and emptying. CRH neurons also stimulate activity of the sympathetic nervous system, and the PVN is considered to be the major integrative site where stressors activate both the HPA axis and the sympathetic nervous system. Some individuals when stressed increase their consumption of "comfort foods," which are generally palatable and calorically dense. Recent evidence suggests that consumption of these foods is a buffer against the magnitude of the physiological responses to both physical and psychological stressors (Ulrich-Lai et al., 2010).

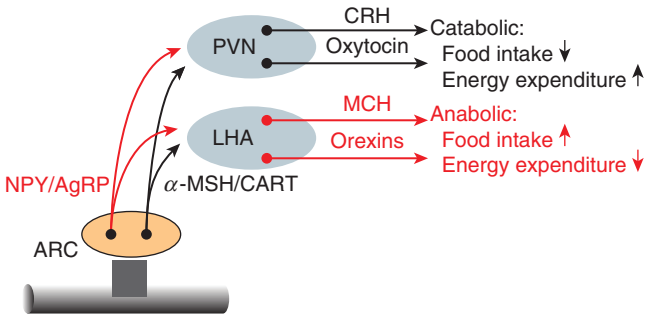


FIGURE 36.6 Hypothalamic neuropeptides that influence caloric homeostasis. The adiposity hormones, leptin and insulin, are transported through the blood–brain barrier and influence neurons in the arcuate nucleus (ARC). ARC neurons that synthesize and release NPY and AgRP are inhibited by adiposity signals, whereas ARC neurons that synthesize and release α-MSH and CART are stimulated by adiposity signals. NPY/AgRP neurons are inhibitory to the PVN and stimulatory to the LHA, whereas α-MSH/CART neurons are stimulatory of the PVN and inhibitory of the LHA. The PVN in turn has a net catabolic action, whereas the LHA has a net anabolic action.

The Brainstem Plays an Important Role in the Control of Food Intake

The NST and the adjacent area postrema in the brainstem receive sensory fibers from gustatory receptors in the mouth and throat (see Chapter 24), as well as afferent information from the stomach, small intestine, pancreas, and liver. In the central control of food intake, these linked brainstem sites likely are where sensory input from the viscera is first integrated with input from taste buds. This sensory information is relayed along prominent neural projections from the brainstem to the hypothalamus, amygdala, and other portions of the limbic system, as well as to the thalamus and gustatory cortex. Reciprocal neural connections to the brainstem from these rostral sites allow emotion and cognitive function to influence the control of eating. Complementary and linked efferent elements participate in the central control of the gastrointestinal tract and liver, and of autonomic function generally.

Evidence of the brainstem's control over food intake comes from investigations of the chronic decerebrate rat in which all axonal connections between the caudal brainstem and the forebrain are severed at the midcollicular level of the midbrain (Grigson, Kaplan, Roitman, Norgen, & Grill, 1997). This animal does not seek food or initiate spontaneous meals, but, like an intact animal, it will reflexively swallow liquid food put directly into its mouth via an oral fistula. The decerebrate rat will swallow a sucrose solution when its stomach is empty, such as after a period of food deprivation; however, it will not swallow water or saline when food deprived, nor will it swallow sucrose solution when its stomach is full or when it has received a systemic injection of CCK or hypertonic saline (but instead lets the administered fluid passively drip from its mouth). These experiments suggest that considerable control of food intake exists entirely within the caudal brainstem, which receives signals that arise when ingested food contacts peripheral sensory organs (i.e., taste buds, gastric stretch receptors, intestinal CCK-secreting cells). Nonetheless, decerebrate animals are incapable of receiving input from forebrain

sites that could modify intake based on learning or other experience, on body adiposity, or on time of day.

Other observations provide additional support for an important role of the brainstem in the control of meal size. When the area postrema is ablated, rats eat larger meals than normal, although total daily food intake is normal; that is, they compensate for their large meals by eating less frequently, indicating that long-term controls of food intake are not impaired. A similar effect is seen in rats pretreated with capsaicin, a neurotoxin that destroys most gastric afferent vagal axons, eliminating the food intake-reducing effects of CCK and presumably damaging the signal of gastric distension.

Summary

We began this chapter with a description of the physiology of caloric homeostasis in order to provide a biological context within which to consider the purpose of eating and the fate of ingested food. Two principles were emphasized. First, meal size is determined by the integrated effects of several acute stimuli. Delivery of calories to the stomach and small intestine upon food consumption and postabsorptive delivery of nutrients to the liver elicit sensory neural signals from the stomach and the liver, respectively, to the brainstem. Second, food intake is also influenced by chronic signals associated with body adiposity. Adiposity indirectly affects gastric and hepatic signals of satiety. In addition, humoral signals proportional to adiposity enter the brain and affect food intake. One signal, leptin, emanates from adipose tissue itself, and the other, insulin, comes from the pancreas. These two peptides appear to be important in the central control of food intake; they modulate the response of the brain to the acute neural signals that affect food intake.

These neural and humoral signals are related to caloric homeostasis and are inhibitory in nature. During meals, they combine to stop ongoing ingestion; later, when these signals disappear in the postabsorptive state, new meals begin. Other inhibitory signals are unrelated to caloric homeostasis yet can have a pronounced effect on the cessation of eating. These signals include gastric distension, toxins (in the food and consumed inadvertently, or else injected systemically), and dehydration.

The central nervous system exerts control over eating at many levels. The spinal cord and brainstem influence all aspects of caloric homeostasis via the autonomic nervous system. The hypothalamus and limbic forebrain receive signals about ingested food and body adiposity and integrate them with information about the taste of the food, the memory of that food, the experience of previous meals, the competition of other desires, and aspects of the environment. Thus, the central control of eating involves many areas of the central nervous system in the collective maintenance of caloric homeostasis.

References

Coleman, D. L., & Hummel, K. P. (1969). Effects of parabiosis of normal with genetically diabetic mice. *American Journal of Physiological, 217,* 1298–1304.

Cone, R. D. (1999). The central melanocortin system and energy homeostasis. *Trends in Endocrinology Metabolism, 10,* 211–216.

Gibbs, J., Young, R. C., & Smith, G. P. (1973). Cholecystokinin decreases food intake in rats. *Journal of Comparative and Physiological Psychology, 84,* 488–495.

Grigson, P. S., Kaplan, J. M., Roitman, M. F., Norgen, R., & Grill, H. J. (1997). Reward comparison in chronic decerebrate rats. *American Journal of Physiological, 273,* R479–R486.

Hagan, M. M., Rushing, P. A., Pritchard, L. M., Schwartz, M. W., Strack, A. M., Van Der Ploeg, L. H., Woods, S. C., & Seeley, R. J. (2000). Long-term orexigenic effects of AgRP-(83–132) involve mechanisms other than melanocortin receptor blockade. *American Journal of Physiological, 279,* R47–R52.

Le Magnen, J., & Tallon, S. (1966). La periodicite spontanee de la prise d'aliments ad libitum du rat blanc. *Journal of Physiological, (Paris), 58,* 323–349.

Leibowitz, S. F. (1990). Hypothalamic neuropeptide Y, galanin, and amines: Concepts of coexistence in relation to feeding behavior. *Annals of the New York Academy of Science, 611,* 221–235.

Moran, T. H. (2010). Hypothalamic nutrient sensing and energy balance. *Forum of Nutrition, 63,* 94–101.

Phillips, R. J. & Powley, T. L. (1996). Gastric volume rather than nutrient content inhibits food intake. *American Journal of Physiological, 271,* R766–R779.

Schwartz, M. W., Woods, S. C., Porte, D., Jr., Seeley, R. J., & Baskin, D. G. (2000). Central nervous system control of food intake. *Nature, 404,* 661–671.

Seeley, R. J., Yagaloff, K. A., Fisher, S. L., Burn, P., Thiele, T. E., van Dijk, G., Baskin, D. G., & Schwartz, M. W. (1997). Melanocortin receptors in leptin effects. *Nature, 390,* 349.

Shimada, M., Tritos, N. A., Lowell, B. B., Flier, J. S., & Maratos-Flier, E. (1998). Mice lacking melanin-concentrating hormone are hypophagic and lean. *Nature, 396,* 670–674.

Smith, G. P., Gibbs, J., & Young, R. C. (1974). Cholecystokinin and intestinal satiety in the rat. *Federation Proceedings, 33,* 1146–1150.

Ulrich-Lai, Y. M., Christiansen, A. M., Ostrander, M. M., Jones, A. A., Jones, K. R., Choi, D. C., Krause, E. G., Evanson, N. K., Furay, A. R., Davis, J. F., Solomon, M. B., De Kloet, A. D., Tamashiro, K. L., Sakai, R. R., Seeley, R. J., Woods, S. C., & Herman, J. P. (2010). Pleasurable behaviours reduce stress via brain reward pathways. *Proceedings of the National Academy of Science of United States of America, 107,* 20529–20534.

Ungerstedt, U. (1971). Adipsia and aphagia after 6-hydroxydopamine induced degeneration of the nigro-striatal dopamine system. *Acta Physiologic Scandinavica Supplementum, 367,* 95–122.

Suggested Readings

Barsh, G. S., Farooqi, I. S., & O'Rahilly, S. (2000). Genetics of bodyweight regulation. *Nature, 404,* 644–651.

Elmquist, J. K., Elias, C. F., & Saper, C. B. (1999). From lesions to leptin: Hypothalamic control of food intake and body weight. *Neuron, 22,* 221–232.

Langhans, W. (1996). Metabolic and glucostatic control of feeding. *The Proceedings of the Nutrition Society, 55,* 497–515.

Levin, B. E., Magnan, C., Dunn-Meynell, A., & Le Foll, C. (2011). Metabolic sensing and the brain: Who, what, where, and how? *Endocrinology, 152,* 2552–2557.

Woods, S. C. (1995). Insulin and the brain: A mutual dependency. *Progress in Psychobiology and Physiological Psychology, 16,* 53–81.

Woods, S. C., Schwartz, M. W., Baskin, D. G., & Seeley, R. S. (2000). Food intake and the regulation of body weight. *Annal Review in Psychology, 51,* 255–277.

Stephen C. Woods and Edward M. Stricker

CHAPTER

37

Water and Salt Intake and Body Fluid Homeostasis

Body fluids are the watery matrix in which the biochemical reactions of cellular metabolism occur. The concentration of substrates in cellular fluid is a key factor in determining the rate at which those reactions take place. All tissues depend on circulating blood to deliver the nutrients needed to support cellular metabolism and to carry away unwanted metabolites for excretion. Thus, the maintenance of solute concentrations or osmolalities—*osmotic homeostasis*—and the regulation of plasma volume—*volume homeostasis*—are essential functions in the physiology of animals.

When normal body fluid osmolality or plasma volume is threatened, various physiological and behavioral responses are stimulated to maintain or restore the basal state adaptively. For example, during water deprivation, animals decrease water lost in urine to prevent dehydration from worsening and they consume water to replace the fluid they have lost. Similarly, hemorrhage stimulates the urinary conservation of water and sodium as well as the ingestion of water and NaCl. Water retention and sodium retention are accomplished through actions of the antidiuretic hormone *arginine vasopressin* (AVP) and the antinatriuretic hormone *aldosterone*, whereas water ingestion and NaCl ingestion are motivated by *thirst* and *salt appetite*. These complementary responses are mediated and coordinated by the brain.

This chapter describes the various mechanisms by which the signals for fluid homeostasis are detected and integrated by the central nervous system. However, we first present a brief overview of body fluid physiology to provide a context in which to consider the regulated functions.

BODY FLUID PHYSIOLOGY

Water is the largest constituent of the body. It contributes 55 to 65% of the body weight of animals, including humans, varying mostly in relation to the amount of body fat. Total body water is distributed between *intracellular fluid* (ICF) and *extracellular fluid* (ECF) *compartments*, with approximately two-thirds in the former and one-third in the latter. The ECF can be further subdivided into the *interstitial fluid* surrounding the cells and the intravascular fluid within blood vessels. The *intravascular fluid*, the plasma (or serum) of blood, averages 7 to 8% of total body water, or 20 to 25% of the ECF.

Fluid compartments differ not only in their volumes but also in the solutes they contain. Specifically, membrane-bound Na^+-K^+ pumps move Na^+ outside the cells and K^+ inside. Despite the differences in solute composition, the *osmotic pressure*, which reflects the concentrations of all solutes in a fluid compartment, is equivalent between ECF and ICF compartments. This equilibrium occurs because water is driven across cell membranes by osmosis from a relatively dilute compartment into one with a higher solute concentration until the osmotic pressures are equivalent on both sides of the cell membrane.

Multiple Mechanisms Help Maintain Blood Volume and Pressure

The distribution of fluid between intravascular and interstitial fluid compartments is determined by a balance between the *hydrostatic pressure* of the blood, which

© 2013 Elsevier Inc. All rights reserved.

is maintained by cardiac output and arteriolar vasoconstriction, and the opposing osmotic pressure contributed by plasma proteins. Although those proteins contribute only 1 to 2% to overall plasma osmolality, the permeability of capillary membranes to such large molecules is low; therefore, proteins exert a pressure differential (of approximately 15–20 mm Hg), called the *colloid osmotic (oncotic) pressure*, that tends to pull interstitial fluid into the circulation. Figure 37.1 summarizes the forces governing transcapillary fluid transfer between the two extracellular compartments, a phenomenon first described by the English physiologist Starling (1896). Note that according to this arrangement, interstitial fluid acts as a reservoir for plasma. The Starling equilibrium is such that when blood pressure falls, as after hemorrhage, interstitial fluid moves into the circulation, thereby helping restore plasma volume. The reverse occurs when saline is added to the blood because plasma proteins are then diluted and the oncotic pressure they provide diminishes, allowing fluid to flow into the interstitial space.

Blood pressure also is maintained by two other mechanisms intrinsic to the cardiovascular system. First, although the arteries that receive the cardiac output of blood have thick walls to preserve blood pressure, the veins are thin-walled, distensible vessels. After a moderate hemorrhage, veins collapse, redistributing blood to arteries that cannot collapse and thus remain full. Consequently, arterial blood pressure is not compromised and the blood deficit occurs primarily on the venous side of the circulation. Conversely, fluid accumulates in the veins when blood volume is expanded, again without much effect on arterial blood pressure. This property is called the *capacitance* or *compliance* of the vascular system. Second, the filtration of blood in the glomeruli of the kidneys is determined in large part by blood pressure in the renal arteries. A drop in blood pressure reduces the *glomerular filtration rate* (GFR) and decreases urine formation, whereas a rise in blood pressure elevates GFR and promotes urinary fluid loss. This normal function of the kidneys is so efficient that the development of hypertension always implicates renal dysfunction as a contributing factor because the kidneys failed to adjust to the elevated blood pressure by increasing fluid excretion in urine.

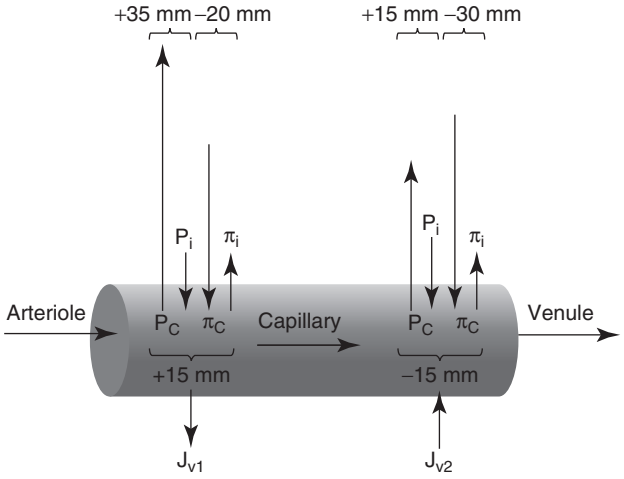


FIGURE 37.1 Starling forces governing transcapillary fluid transfer. At the arteriolar end of the capillary, the difference between the intravascular hydrostatic pressure (P_c) and the interstitial hydrostatic pressure (P_i) exceeds the oppositely oriented difference between the intravascular oncotic pressure (π_c) and the interstitial oncotic pressure (π_i); the resultant pressure gradient drives capillary fluid into the interstitial space (J_{v1}). As fluid leaves the capillary, P_c decreases due to fluid loss and π_c increases due to hemoconcentration. Consequently, at the venous end of the capillary, interstitial fluid is pulled back into the vascular space (J_{V2}). Numerical values indicate approximate net pressure differences (in mm Hg) between intravascular and interstitial spaces. Relative sizes of P and π are indicated by arrow length. Note that fluid accumulating in the interstitial space ultimately returns to the blood via the lymphatic system (not shown).

Summary

Body fluid homeostasis is directed at achieving stability in the osmolality of body fluids and the volume of plasma. Such homeostatic regulation is promoted by several mechanisms intrinsic to the physiology of body fluids and the cardiovascular system. Nevertheless, changes in body fluid osmolality and plasma volume may be so large that additional mechanisms must be recruited to maintain homeostasis. These other responses involve the central control of water and sodium excretion in urine through the actions of specific hormones, and the central control of water and NaCl consumption motivated by thirst and salt appetite.

OSMOTIC HOMEOSTASIS

Osmolality is an expression of concentration—that is, a ratio of the total amount of solute dissolved in a given weight of water:

$$\frac{\text{solute (osmoles)}}{\text{water (kilograms)}}.$$

Dehydration and the consequent need for water occur whenever this ratio is elevated, whether by a decrease in its denominator or by an increase in its numerator. In fact, both changes occur naturally and often: Body water decreases as a result of water deprivation or the loss of dilute fluids to accomplish evaporative cooling (e.g., sweating), and it increases as a result of solute load (e.g., the consumption of NaCl in foods).

The water loss associated with dehydration is borne by the ECF and the ICF in proportion to their sizes because the osmolality of fluid in the two compartments remains in equilibrium. In contrast, the increase in plasma osmolality that results from a solute load results only in cellular dehydration because the water leaving cells by osmosis expands ECF volume. Thus, a solute load is a more abrupt, less complex treatment than water deprivation for stimulating AVP secretion and thirst, the two main osmoregulatory responses of the brain.

Arginine Vasopressin Is the Antidiuretic Hormone

AVP is a nine amino acid peptide that is synthesized in the magnocellular neurons in the supraoptic nucleus (SON) and the paraventricular nucleus (PVN) of the hypothalamus. The peptide is cleaved enzymatically from its prohormone and transported along axons projecting from the hypothalamus to the nearby posterior lobe of the pituitary gland (*neurohypophysis*). There, AVP is stored within neurosecretory granules until specific stimuli, such as an increase in the effective osmolality of body fluids, activate its secretion into the bloodstream. The importance of AVP in maintaining water balance is underscored by the fact that its stores in the pituitary contain sufficient hormone to enable maximal antidiuresis when dehydration is sustained for more than a week.

The circulating hormone acts on a subset of AVP receptors (the V2 subtype) in the kidney to increase water permeability of the collecting ducts of nephrons through insertion of water channels, called aquaporins, into the apical membranes of tubular epithelial cells. Antidiuresis occurs when water moves out of the distal convoluted tubule and collecting duct by osmosis. As secondary responses to the increased net water reabsorption, urine flow decreases and urine osmolality increases. See Box 37.1 for more on renal water channels.

With refinement of radioimmunoassays for AVP, the unique sensitivity of the hormone to small changes in osmolality has become apparent, as has the remarkable sensitivity of the kidney to small changes in plasma AVP levels. Circulating AVP is linearly related to plasma osmolality above a threshold of 1 to 2% (Fig. 37.2). Urine osmolality, in turn, is related linearly to AVP levels from 0.5 to 5–6 $pgml^{-1}$, in association with increases in plasma osmolality to only 4% above the threshold for AVP secretion (Fig. 37.3). However, because urine volume is related inversely to urine osmolality, small increases in plasma AVP concentration (e.g., from 0.5 to 2 $pgml^{-1}$) have the effect of decreasing urine flow much more than subsequent larger increases (e.g., from 2 to 5 $pgml^{-1}$; Fig. 37.3). This relationship emphasizes the marked physiological effects of small initial changes in plasma AVP levels. The net result of these relations among plasma osmolality, AVP secretion, urine volume, and urine osmolality is a finely tuned regulatory system that adjusts the rate of free water excretion according to plasma osmolality via changes in pituitary hormone secretion.

Thirst Is Another Effective Osmoregulatory Response to Dehydration

Accompanying the excretion of concentrated urine is reabsorption of conserved water, which can considerably dilute remaining body fluids. Thirst, and the water intake it provokes, is a much more rapid and less limited response to dehydration than antidiuresis. Thirst may be defined as a strong motivation to seek, to obtain, and to consume water in response to deficits in body fluids. Like AVP secretion, thirst can be stimulated by cellular dehydration caused by increases in the effective osmolality of ECF (Gilman, 1937). Studies in animals consistently indicate that drinking behavior is elicited by 1 to 3% increases in plasma osmolality above basal levels, and analogous research has revealed that similar thresholds must be reached to produce thirst in humans (Robertson, 1986). This arrangement, in which the threshold for drinking is slightly higher than that for secretion of AVP (Fig. 37.3), ensures that ongoing behavior is not disrupted by thirst unless the buffering effects of osmosis and antidiuresis are insufficient to maintain osmoregulation. It also ensures that dehydration does not become severe before thirst is stimulated. See Box 37.2 for more on how the neurohypophysis controls the volume of water in the body.

Also like AVP secretion, water intake increases linearly in proportion to increases in the effective osmolality of ECF. The dilution of body fluids by ingested water complements the retention of water that occurs during antidiuresis, and both responses occur when drinking water is available. However, there may be marked individual differences in whether dehydrated subjects respond to their need for water promptly by drinking or more slowly by increasing renal water conservation (Kanter, 1953). AVP secretion and urine osmolality are more elevated when an induced increase in plasma osmolality is not compensated for by water intake (e.g., because drinking water is not available). Conversely, water intake in response to a solute load increases in animals after their kidneys are removed or their capability to secrete AVP is compromised, thereby precluding rapid excretion of the load in concentrated urine.

BOX 37.1

WATER CHANNELS

Water conservation by the kidney is dependent on its ability to concentrate the urine. Urine is concentrated through the combined actions of the loop of Henle and the collecting duct of the kidney. The loop of Henle generates a high osmolality in the renal medulla via a mechanism known as the countercurrent multiplier system. AVP acts in the collecting duct to increase water permeability, thereby allowing osmotic equilibration between the urine and the hypertonic interstitium of the renal medulla. The net effect of this process is to extract water from the urine into the medullary interstitial blood vessels, resulting in increased urine concentration and decreased urine volume.

AVP produces antidiuresis by virtue of its effects on the epithelial principal cells of the collecting duct in the kidney, which are endowed with AVP receptors of the V2 type. It has long been known that binding of AVP to G protein-coupled V2 receptors causes cAMP generation via activation of adenylate cyclase. However, the intracellular mechanisms responsible for the subsequent increased water reabsorption across the collecting duct cells were elucidated only after the discovery of *aquaporins*, which are a widely expressed family of water channels that facilitate rapid water transport across cell membranes.

Many different aquaporin water channels are expressed in various body tissues, including the brain. Several of these channels have been localized in the kidney. Aquaporin-1 is expressed constitutively in the proximal tubule and thin descending limb of the loop of Henle and is believed to be responsible for reabsorption of a large fraction of the water filtered by the glomerulus. The other three water channels, aquaporin-2, -3, and -4, are expressed in collecting duct principal cells, target cells for the action of AVP to regulate collecting duct water permeability. Aquaporin-2 is the only water channel known to be expressed in the apical membrane (i.e., the side bordering the tubular lumen, through which urine flows) of collecting duct principal cells and is also abundant in intracellular vesicles located below the apical membrane. Aquaporin-2 is the major AVP-regulated water channel and mediates water transport across the apical plasma membrane of the principal cells of the collecting ducts. In contrast, aquaporins-3 and -4 are expressed at high levels in the basolateral plasma membranes (i.e., the side bordering the blood) of principal cells and are responsible for the constitutively high water permeability of the basolateral plasma membrane.

There are at least two ways by which AVP regulates osmotic water permeability in the collecting duct: short-term and long-term regulation (Knepper, 1997). Short-term regulation is associated with increases in water permeability within a few minutes of AVP exposure, an effect that is rapidly reversible. AVP triggers this response by binding to the V2 receptor and increasing intracellular cyclic AMP levels by activating adenylate cyclase. The increase in collecting duct water permeability is a consequence of fusion of aquaporin-2-containing intracytoplasmic vesicles with the apical plasma membranes of the principal cells, a process that increases apical water permeability by markedly increasing the number of water-conducting pores in the apical plasma membrane. Dissociation of AVP from the V2 receptor allows intracellular cAMP levels to decrease, and the water channels are reinternalized into the intracytoplasmic vesicles, thereby terminating the increased water permeability. By virtue of the subapical membrane localization of the aquaporin-containing vesicles, they can be quickly shuttled into and out of the membrane in response to changes in intracellular cAMP levels. This mechanism therefore allows minute-to-minute regulation of renal water excretion through changes in ambient plasma AVP levels. Long-term regulation of collecting duct water permeability represents a sustained increase in collecting duct water permeability in response to prolonged high levels of circulating AVP. This response requires at least 24 h to elicit and is not as rapidly reversible. This conditioning effect is due largely to the ability of AVP to induce large increases in the abundance of aquaporin-2 and -3 water channels in the collecting duct principal cells. Greater total expression of the number of aquaporin-2 and -3 water channels, when combined with the short-term effect of AVP to shift aquaporin-2 into the apical plasma membrane, allows the collecting ducts to achieve extremely high water permeabilities during conditions of prolonged dehydration, thereby further enhancing the urine concentrating capacity in response to elevated levels of circulating AVP.

Rapid Inhibitory Feedback Control of Drinking

When dehydrated, many animals drink water very quickly, stopping long before the ingested water actually rehydrates them. This anticipatory satiety appears to be mediated by a neural message communicated from the oropharynx to the brain in association with the act of swallowing. The main experiments were done in dogs with implanted gastric fistulas (Thrasher, Nistal-Herrera, Keil, & Ramsay, 1981). When the fistula was

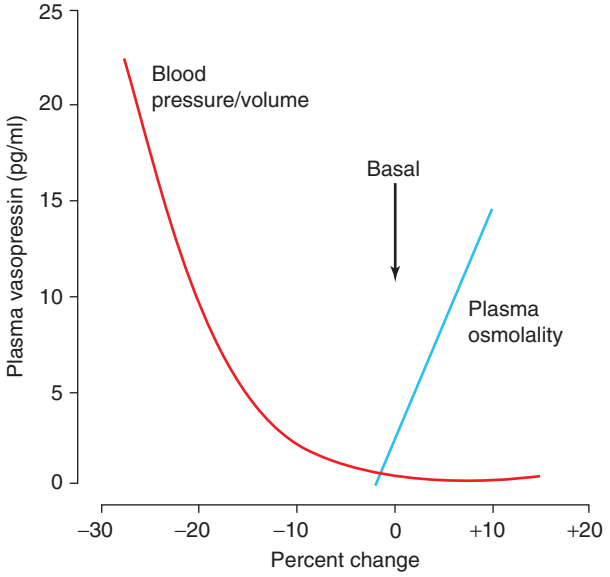


FIGURE 37.2 Plasma concentrations of AVP as a function of changes in plasma osmolality, blood volume, or blood pressure in humans. The arrow indicates the plasma AVP concentration at basal plasma osmolality, volume, and blood pressure. *Modified with permission from Robertson (1986).*

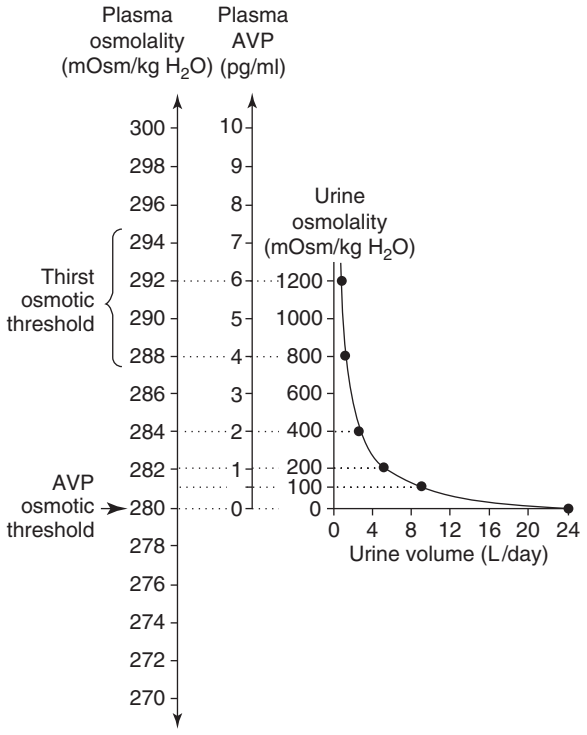


FIGURE 37.3 Relationship of plasma osmolality, plasma AVP concentrations, urine osmolality, and urine volume in humans. Note that small changes in plasma AVP concentrations have larger effects on urine volume at low plasma AVP concentrations than at high plasma AVP concentrations. *Modified with permission from Robinson (1985).*

closed, ingested water flowed normally from the stomach into the small intestine and was absorbed; water-deprived dogs drank rapidly, stopped 15 min before their plasma osmolalities began to decrease, and did not resume drinking as their stomach emptied because plasma osmolalities had returned to normal. In contrast, when the fistula was left open, ingested water drained out from the stomach, never entering the intestine and thus never being absorbed; the dogs drank normal amounts, stopped drinking, and after 10 to 20 min resumed drinking. In another study, when the dogs were not allowed to drink but water was introduced into the stomach through the fistula, thus bypassing the oropharynx, the gastric load did not affect water intake until it had been absorbed.

These seminal observations are consistent with a temporary satiety being produced by the act of water consumption, with a more long-lasting satiety occurring only after absorption of the ingested water and rehydration of the dog. AVP secretion is affected in parallel, showing that physiological and behavioral components of osmoregulation are influenced similarly. Ingestion of a concentrated NaCl solution instead of water produced the same temporary satiety and inhibition of AVP secretion, but after the saline was absorbed, the increase in plasma osmolality stimulated more drinking and AVP secretion than occurred at first. In similar experiments, people responded the same way but rats appear to have a different mechanism of rapid feedback, perhaps involving peripheral osmoreceptors.

Osmoreceptor Cells Stimulate AVP Secretion and Thirst

All body cells lose water by osmosis when the effective osmolality of ECF is increased. Thus, cells that provoke AVP secretion and thirst do not have unique osmosensitive properties (unlike retinal photoreceptor cells, for example, which are uniquely responsive to light). Instead, the exceptional feature of osmoreceptor cells is thought to be their neural circuitry, which activates the central systems for AVP secretion and thirst when the cells are dehydrated.

Destruction of osmoreceptor neurons should eliminate detection of increased plasma osmolality and thus the AVP secretion and thirst responses that are elicited by dehydration. In fact, ample research has confirmed that certain brain lesions eliminate AVP secretion and thirst responses. Such studies also have revealed that osmoreceptor cells (Johnson & Buggy, 1978; Thrasher, Keil, & Ramsay, 1982) appear to be located in the vascular organ of the lamina terminalis (OVLT) and areas of the adjacent anterior hypothalamus, near the anterior wall of the third cerebral ventricle (Fig. 37.4). Surgical destruction of that

BOX 37.2

DIABETES INSIPIDUS

The disease *diabetes insipidus* (DI), in which the secretion of AVP is impaired or absent, illustrates the crucial role of AVP in controlling the volume of water in the body. An eighteenth century Scottish physician, William Cullen, named the disease diabetes insipidus, or insipid urine, to distinguish it from *diabetes mellitus*, or sweet urine, in which abnormally high concentrations of glucose in the blood result in glucose appearing in the urine.

AVP is the only known antidiuretic substance in the body. In its absence, the kidney is unable to concentrate urine maximally to conserve water. The result is a continued excretion of copious amounts of very dilute urine. Patients with severe cases of DI, in whom the ability to excrete AVP is completely lost, can excrete up to 25 liters of urine each day. Such patients urinate almost hourly, which renders the completion of even simple tasks and activities of daily living, including sleeping, exceedingly difficult.

Patients with DI can quickly become dehydrated if their urinary fluid losses are not replaced by drinking water. Fortunately, thirst remains intact in most patients with DI because lesions that destroy the magnocellular neurons in the SON and PVN that synthesize AVP generally leave intact the osmoreceptors in the anterior hypothalamus as well as the higher brain centers that control thirst. Consequently, extreme thirst is one of the hallmarks of this disease, leading to the characteristic symptoms of *polydipsia* (excessive drinking) and *polyuria* (excessive urination). If drinking water is unavailable, or if a person with DI is unable to drink, then the unreplaced urinary water loss leads to dehydration and death in the absence of medical intervention.

DI can be caused when tumors and infiltrative diseases of the hypothalamus destroy the magnocellular neurons that produce AVP. Because four nuclei contain magnocellular neurons, and 10 to 20% of AVP-producing neurons are sufficient for normal urine concentration, brain lesions that cause DI are generally large. Less commonly, DI is idiopathic (of unknown cause), most likely from an autoimmune basis. DI can also be genetic, transmitted as an autosomal-dominant trait. Some patients with DI do not have any defect in AVP secretion but rather have defects in the V2 AVP receptors in the kidney that respond to circulating AVP. These cases are called *nephrogenic* (of kidney origin) DI.

The treatment for DI of hypothalamic origin, like that of other endocrine deficiency disorders, is replacement of the deficient hormone, in this case AVP. The short half-life of AVP in the circulation allows mammals to have minute-to-minute control of their urine output. However, longer-acting synthetic analogs of AVP are more convenient for treatment because they need not be taken as frequently as short-acting drugs (Robinson, 1985). These agents can restore urinary concentration and allow a person with DI to lead a more normal life.

area of the brain in animals abolishes the AVP secretion and thirst responses to hyperosmolality but not their responses to other stimuli. The same conclusion was drawn after a study of people who were unable to osmoregulate when water deprived or when given an NaCl load; these patients were found to have focal brain tumors that destroyed the region around the OVLT.

The location of osmoreceptor cells in the OVLT is consistent with the results of pioneering investigations in which hyperosmotic solutions injected into blood vessels perfusing the anterior hypothalamus stimulated AVP secretion in dogs (Verney, 1947). The OVLT and surrounding areas of the anterior hypothalamus were subsequently implicated by studies in rats. After systemic injections of hypertonic NaCl solution, modern immunocytochemical techniques were used to detect early gene products in cells associated with the production of new protein. Dense staining in the OVLT (and in AVP-secreting cells in the hypothalamus) indicated that it had been strongly stimulated by the experimentally induced dehydration and closely paralleled pituitary AVP secretion (Hoffman, Smith, & Verbalis, 1993).

The neural pathways connecting the OVLT with magnocellular AVP-secreting cells in the hypothalamic SON and PVN have been identified, whereas the neural circuits in the forebrain that control thirst are less well known. Early reports identified the lateral hypothalamus as a "thirst center" because its destruction in rats eliminated the drinking response, but not the associated AVP secretion, to increased osmolality of body fluids. However, later investigations showed that the critical damage was not to hypothalamic cells but to dopamine-containing fibers that coursed through the area. The induced disruption of behavior was not specific to drinking but instead reflected a general inability of the brain-damaged animals to initiate movement (Stricker, 1976). Indeed, the syndrome of behavioral dysfunctions seen in these animals generally resembled that of Parkinson's disease, which has also been attributed to the loss of dopaminergic fibers in the brain. (See Chapter 36 for the same reinterpretation

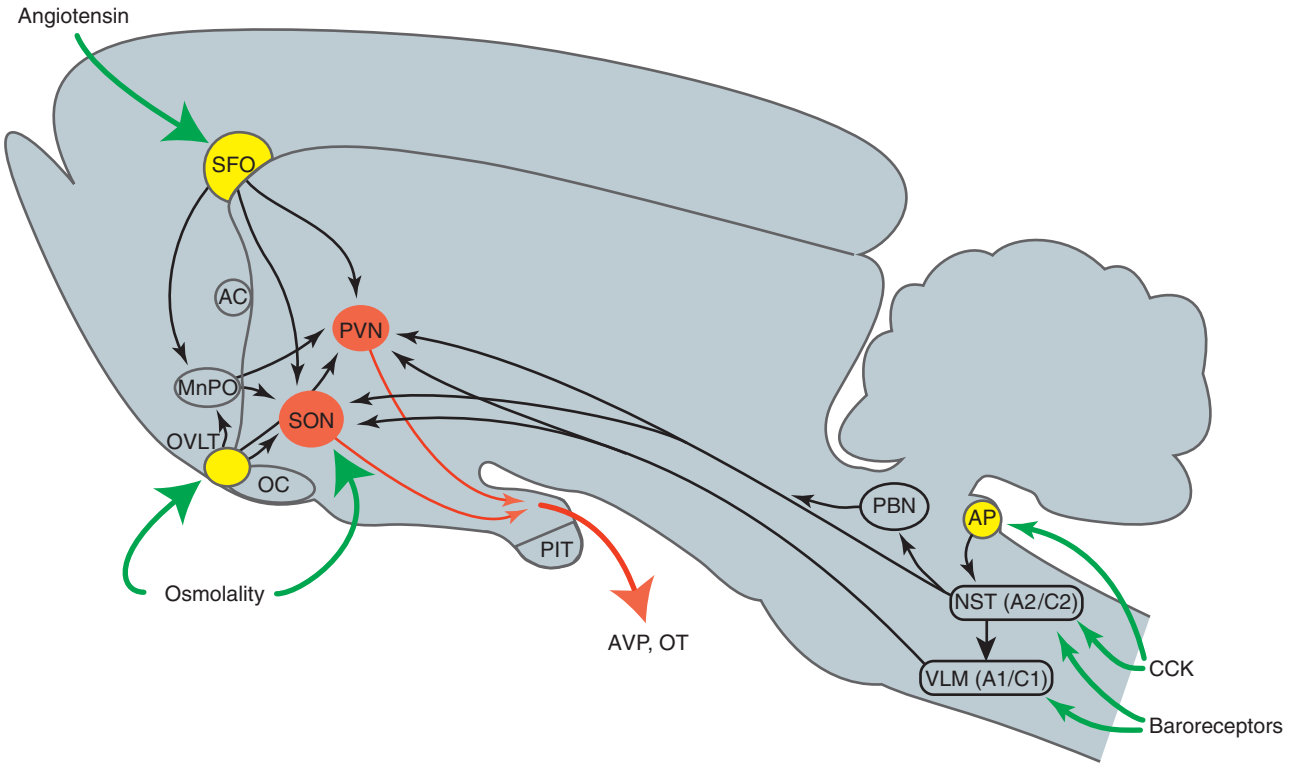


FIGURE 37.4 Summary of the main anterior hypothalamic pathways that mediate secretion of arginine vasopressin (AVP) and oxytocin (OT). The vascular organ of the lamina terminalis (OVLT) is especially sensitive to hyperosmolality. Hyperosmolality also activates other neurons in the anterior hypothalamus, such as those in the subfornical organ (SFO) and median preoptic nucleus (MnPO), and magnocellular neurons, which are intrinsically osmosensitive. Circulating angiotensin II (AII) activates neurons of the SFO, an essential site of AII action, as well as cells throughout the lamina terminalis and MnPO. In response to hyperosmolality or AII, projections from the SFO and OVLT to the MnPO activate excitatory and inhibitory interneurons that project to the supraoptic nucleus (SON) and paraventricular nucleus (PVN) to modulate direct inputs to these areas from the circumventricular organs. Cholecystokinin (CCK) acts primarily on gastric vagal afferents that terminate in the nucleus of the solitary tract (NST), but at higher doses it can also act at the area postrema (AP). Although neurons apparently are activated in the ventrolateral medulla (VLM) and NST, most oxytocin secretion appears to be stimulated by monosynaptic projections from A2/C2 cells, and possibly also noncatecholaminergic somatostatin/inhibin b cells, of the NST. Baroreceptor-mediated stimuli, such as hypovolemia and hypotension, are more complex. The major projection to magnocellular AVP neurons appears to arise from A1 cells of the VLM that are activated by excitatory interneurons from the NST. Other areas, such as the parabrachial nucleus (PBN), may contribute multisynaptic projections. Cranial nerves IX and X, which terminate in the NST, also contribute input to magnocellular AVP neurons. It is unclear whether baroreceptor-mediated secretion of oxytocin results from projections from VLM neurons or from NST neurons. AC, anterior commissure; OC, optic chiasm; PIT, anterior pituitary.

of the inability of rats to eat normally after dopamine-depleting lesions of the lateral hypothalamus.) More recent studies using functional magnetic resonance imaging of the brain have implicated the anterior cingulate cortex as a likely area involved in thirst perception (Egan et al., 2003).

Natriuresis and Inhibition of Solute Intake Also Promote Osmotic Homeostasis during Dehydration

When body fluid is hyperosmolar, adaptive behavior includes not only drinking and conserving water (thereby increasing the denominator in the ratio that represents body fluid osmolality) but also excreting NaCl and avoiding the consumption of additional osmolytes (thereby decreasing the numerator of that ratio). Endogenous natriuretic agents promote urinary sodium loss after an administered NaCl load or a period of imposed water deprivation. One such agent is the hormone *atrial natriuretic peptide* (ANP), which is synthesized in the atria of the heart and released when increased intravascular volume distends the atria. Another is the hormone *oxytocin*. Like AVP, oxytocin is synthesized in magnocellular neurons in the PVN and SON and secreted from the posterior pituitary in proportion to induced hyperosmolality (Stricker & Verbalis, 1986). In rats, oxytocin is as potent in stimulating natriuresis as AVP is in stimulating antidiuresis (Verbalis, Mangione, & Stricker, 1991), by acting directly in the kidneys and indirectly by eliciting secretion of ANP.

Salt loads are also known to decrease the intake of osmolytes, whether in the form of NaCl solution or

food, complementing the stimulation of thirst and the secretion of AVP and oxytocin. However, because destruction of the OVLT eliminates the two latter effects but not the dehydration-induced reduction in NaCl intake, osmo- or Na^+ receptors located outside the basal forebrain must mediate the inhibition of NaCl and food intake. Possible sites for such cells include the *hepatic portal vein* and the *area postrema*, both of which have been suspected of having Na^+ receptor functions. In addition, cells in the hepatic portal vein are well situated to detect the sodium content of ingested food and to modulate its intake accordingly.

Diuresis and Inhibition of Water Intake Promote Osmotic Homeostasis during Overhydration

Osmoregulation is required not only under conditions of dehydration but also during periods of acute overhydration and hypoosmolality, as may result when beverages are consumed in excess of water needs or when AVP levels are non-osmotically elevated. Such consumption may occur not because of thirst but, for example, because of the palatability of or chemical substances in the beverages (e.g., caffeine, alcohol). (See Box 37.3 on water intoxication.) Unlike excess food, which is stored as triglycerides in adipose tissue, excess water is not stored for later use but instead is excreted in urine. When plasma osmolality is below normal, circulating levels of AVP are reduced and, in consequence, the kidneys excrete dilute urine and thereby raise plasma osmolality. The major behavioral contribution of osmotic dilution to osmoregulation is inhibition of free water intake; the ingestion of osmolytes in food and NaCl is not stimulated by osmotic dilution.

Summary

Osmoregulation during dehydration in animals and humans is accomplished by a combination of physiological responses, resulting in antidiuresis and natriuresis, and the behavioral response of increased water intake. Osmoreceptor cells critical for mediating these functions have been identified in the basal forebrain. These neurons respond to very small increases in plasma osmolality, and the effector systems they control correct any increase in plasma osmolality. In addition, an early stimulus must exist, generated by peripheral osmo- or Na^+ receptor cells, which signals the brain in anticipation of subsequent rehydration. Other osmo- or Na^+ receptor cells appear to mediate the inhibition of NaCl and food intake, which also contributes to osmoregulation during dehydration. Conversely, diuresis and inhibition of water intake promote osmoregulation during overhydration, whereas excitation of NaCl or food intake does not.

VOLUME HOMEOSTASIS

Like osmotic dehydration, loss of blood volume (hypovolemia) stimulates several adaptive compensatory responses appropriate for restoring circulatory volume. The physiological contributions to volume regulation have been studied extensively in laboratory animals subjected to a controlled loss of blood. Because behavior is compromised by the anemia and hypotension that result from extensive hemorrhage, researchers studying thirst in rats often instead produce hypovolemia by subcutaneous injection of a colloidal solution. Such treatment disrupts the Starling equilibrium in capillaries near the injection site because the extravascular colloid opposes the oncotic effect of plasma proteins. Consequently, fluid leaches out of capillaries and remains in the interstitial space. Ingested fluid is also drawn into the interstitial space by the colloid so the total fluid volume required to correct the plasma volume deficit may be substantial. Water does not move from the cells by osmosis, however, because the osmolality of the injected colloidal solutions actually is similar to that of cells.

After colloid injections, rats conserve water and Na^+ in urine and increase their consumption of water and saline solution (Stricker, 1981). When given an isotonic NaCl solution to drink, these rats ingest volumes appropriate to their needs. When given separate bottles of water and concentrated NaCl solution, remarkably the rats drink appropriate amounts of each to create an isotonic fluid mixture. These observations raise several questions: How do the hypovolemic rats detect their plasma volume depletion? How are their thirst and salt appetite coordinated so that they consume the desired isotonic mixture of fluid? How are these two behavioral responses integrated with the complementary physiological responses of water and Na^+ conservation in urine? Research has provided answers to these and related questions, which are discussed in the following sections.

Neural and Endocrine Signals of Hypovolemia Stimulate AVP Secretion

An appropriate physiological response to volume depletion should include water conservation and urine concentration. In fact, like plasma hyperosmolality, hypovolemia is an effective stimulus for AVP secretion (Robertson, 1986; Stricker & Verbalis, 1986). However, AVP secretion does not occur until blood loss exceeds 10% of total blood volume, meaning AVP secretion is much less sensitive to hypovolemia than to increases in ECF osmolality (Fig. 37.2). The effects of hypovolemia and osmotic dehydration are additive; that is, a given increase in osmolality causes greater secretion of AVP when animals are hypovolemic than when they are euvolemic (Fig. 37.5).

BOX 37.3

WATER INTOXICATION

Overhydration that leads to retention of excess water is commonly called water intoxication. This can be caused by two mechanisms: excessive ingestion of fluids in volumes greater than the ability of the kidneys to excrete water or impaired kidney water excretion. In many cases, both mechanisms contribute to the water retention. Ingestion of dilute fluids normally causes slight decreases in plasma osmolality, which are sufficient to reduce circulating AVP concentrations to levels that allow the kidneys to excrete the ingested water in dilute urine, thereby maintaining osmotic homeostasis. The capacity of mammalian kidneys to excrete excess fluid is very large, in humans ranging from 800 to 1,000 ml/h, or 19 to 24 L/d. Humans rarely ingest fluids at these large volumes; however, this does occur in patients with schizophrenia, who drink compulsively for reasons not related to thirst, or during periods of forced water consumption as has occurred during fraternity hazing or freshwater drowning. More commonly, water retention occurs when the water excreting capacity of the kidney is limited by increased plasma AVP levels. This can occur because of physiological nonosmotic stimuli that cause AVP secretion, such as hypovolemia, or because of pathological AVP secretion caused by a variety of disease processes, including tumors that synthesize and secrete AVP or central nervous system disorders that interfere with normal regulation of AVP secretion, which is called the *syndrome of inappropriate antidiuretic hormone secretion* (SIADH) (Verbalis, 2007).

The metabolic consequence of excess water retention is a dilution of the osmolality of body fluids, or *hypoosmolality*, and a commensurate dilution of the extracellular fluid (ECF) sodium concentration to low levels, or *hyponatremia*. The physiological consequence of lowered ECF osmolality is an imbalance of osmolality across the intracellular fluid (ICF) and ECF compartments. In response, water will move across cell membranes from the ECF to the ICF until osmotic equilibrium is again achieved. This water movement results in swelling of all cells throughout the body. In most organs this is well tolerated, but the brain is enclosed in a rigid bony skull that limits the degree of brain swelling, or *cerebral edema*, to approximately 8%, which is the space occupied by cerebrospinal fluid within the skull. Once this limit is reached, the brain herniates through the openings at the base of the brain, which compresses the brainstem respiratory regulatory center leading to cessation of breathing and death. This pathological process is the cause of death in cases of water intoxication that develops rapidly, as in individuals engaging in endurance exercise events such as marathon races, who ingest too much fluid relative to their reduced urine excretory capacity as a result of exercise-induced stimulation of AVP secretion (Noakes et al., 2005). Individuals with lesser degrees of cerebral edema nonetheless can develop neurological manifestations ranging from disorientation to seizures, obtundation, and coma, a syndrome called *hyponatremic encephalopathy* (Fraser & Arieff, 1997).

However, such dire consequences do not develop in most patients with hypoosmolality and hyponatremia because the brain is able to adapt to hypoosmolality via a process called *brain volume regulation* (Verbalis, 2010). In response to hypoosmolality-induced cell swelling, the brain loses electrolytes and small molecules culled from the ICF and ECF spaces, thereby decreasing brain water content back toward normal levels. This process generally requires 24–48 hours in both experimental animals and humans. The marked variability in the presenting neurological symptoms of hyponatremic patients can be understood in the context provided by this process of brain volume regulation. Most of the severe neurological symptoms associated with hyponatremia are thought to reflect the cerebral edema that occurs as a consequence of osmotic water movement into the brain. However, once the brain has volume adapted through solute losses, thereby reducing the cerebral edema, neurological symptoms are not as prominent and in some cases may even be totally absent. On the other hand, even patients with chronic hyponatremia appear to be at risk from deleterious complications as a result of the cellular solute losses that allow survival during hypoosmolar conditions. These include altered neurocognition, gait instability, increased falls and bone loss leading to bone fractures, and a debilitating demyelinating disorder called central pontine myelinolysis if the hyponatremia is corrected too rapidly.

Loss of blood volume is first detected by stretch receptors in the great veins entering the right atrium of the heart. These stretch receptors provide an afferent vagal signal to the *nucleus of the solitary tract* (NST) in the brainstem (Chapter 36). Still larger decreases in blood volume may also lower arterial blood pressure and reduce the stretch of receptors in the walls of the carotid sinus and aortic arch. That information is integrated in the NST with neural messages from the low-pressure, venous side of the circulation. Note that these sensory neurons are not actually baroreceptors (literally, pressure receptors); although commonly they

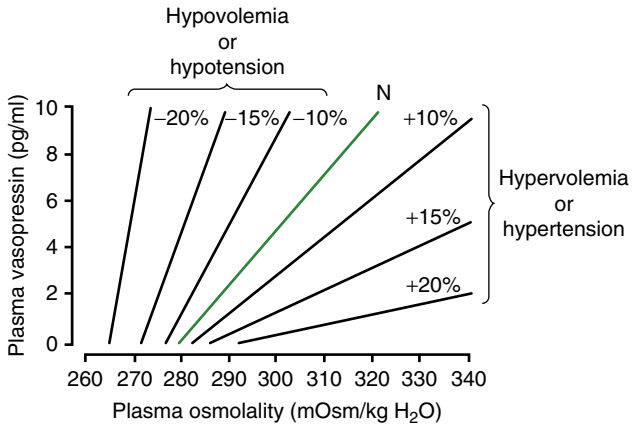


FIGURE 37.5 The relationship between the osmolality of plasma and the concentration of vasopressin (AVP) in plasma depends on blood volume and pressure. The line labeled N shows plasma vasopressin concentration across a range of plasma osmolality in an adult with normal intravascular volume (euvolemic) and normal blood pressure (normotensive). Lines to the left of N show the relationship between plasma vasopressin concentration and plasma osmolality in adults whose low intravascular volume (hypovolemia) or blood pressure (hypotension) is 10, 15, and 20% below normal. Lines to the right of N are for volumes and blood pressures 10, 15, and 20% above normal. *Modified with permission from Robertson (1986).*

are referred to as such, they are more accurately described as stretch receptors.

The ascending pathway between the NST in the brainstem and the SON and PVN in the hypothalamus includes noradrenergic fibers arising from A1 cells in the ventrolateral medulla. Volume depletion stimulates AVP secretion via other pathways as well. In addition, the kidneys secrete renin during hypovolemia, a response mediated in part by sympathetic neural input to β-adrenergic receptors on cells that secrete renin. Renin is an enzyme that initiates a cascade of biochemical steps that result in the formation of *angiotensin II* (AII) (Fig. 37.6), an extremely potent vasoconstrictor. AII stimulates AVP secretion by acting in the brain at the *subfornical organ* (SFO) (Ferguson & Renaud, 1986), which is located in the dorsal portion of the third cerebral ventricle. Because this circumventricular organ lacks a blood–brain barrier, its AII receptors can detect very small increases in the blood levels of AII. Neural pathways from the SFO to the SON and PVN in the hypothalamus mediate AVP secretion and may use AII as a neurotransmitter. Another pathway from the SFO goes to the OVLT, perhaps providing an opportunity for the integration of information about volume states and osmotic concentration.

Neural and Endocrine Signals of Hypovolemia Also Stimulate Thirst

In addition to the antidiuresis and vasoconstriction produced by the secretion of AVP and activation of the renin–angiotensin system, colloid treatment increases water intake in rats (Fitzsimons, 1961). Once 5 to 10% of the normal plasma volume has been lost, the water intake elicited by hypovolemia increases linearly in relation to further deficits in plasma volume. This stimulus for water intake and the effect of an NaCl load on thirst are additive. The stimulus of thirst during hypovolemia appears to be the same as the signal for AVP secretion—that is, a combination of neural afferents from cardiovascular baroreceptors and endocrine stimulation by AII (Fitzsimons, 1969). Each signal can stimulate water intake in the absence of the other. For example, water intake by hypovolemic rats is not diminished by destruction of the NST and loss of neural input from baroreceptors, nor is it eliminated by the loss of AII resulting from bilateral removal of the kidneys (Fitzsimons, 1961). Presumably, colloid treatment would not elicit thirst in rats subjected concurrently to bilateral nephrectomy and NST lesions.

Intravenous infusion of AII strongly stimulates thirst in rats and most other animals studied in the laboratory (including humans). AII adds to the thirst stimulated by an osmotic load when the two treatments are combined. There had been considerable controversy about whether AII functions as a normal physiological stimulus of thirst because the doses of AII required to stimulate significant water intake produce blood levels of AII well above the physiological range. However, such doses also increase arterial blood pressure, and it has been shown that this acute hypertension inhibits thirst and limits the induced water intake. For example, drinking was enhanced considerably when the hypertensive effects of AII were blunted by simultaneous administration of vasodilating agents (Robinson & Evered, 1987).

Osmotic Dilution Inhibits Thirst and AVP Secretion during Hypovolemia

Hypovolemic rats need isotonic saline, not water alone, to repair their plasma volume deficits. When the rats drink only water, about two-thirds of the ingested volume moves into cells by osmosis and much of what remains extracellular is captured by the colloid. Thus, most of the water that is consumed does not remain in the vascular compartment, and the volume deficit within the vascular compartment persists. This situation may be contrasted with the negative feedback control of osmoregulatory thirst in which the dehydration of osmoreceptor cells causes thirst and ingested water repairs dehydration, eliminating the stimulus for thirst.

Rather than repairing the volume deficit, the water consumed by hypovolemic rats causes a second serious challenge to body fluid homeostasis: osmotic dilution. The animals cannot readily eliminate this self-administered water load because hypovolemia reduces GFR and

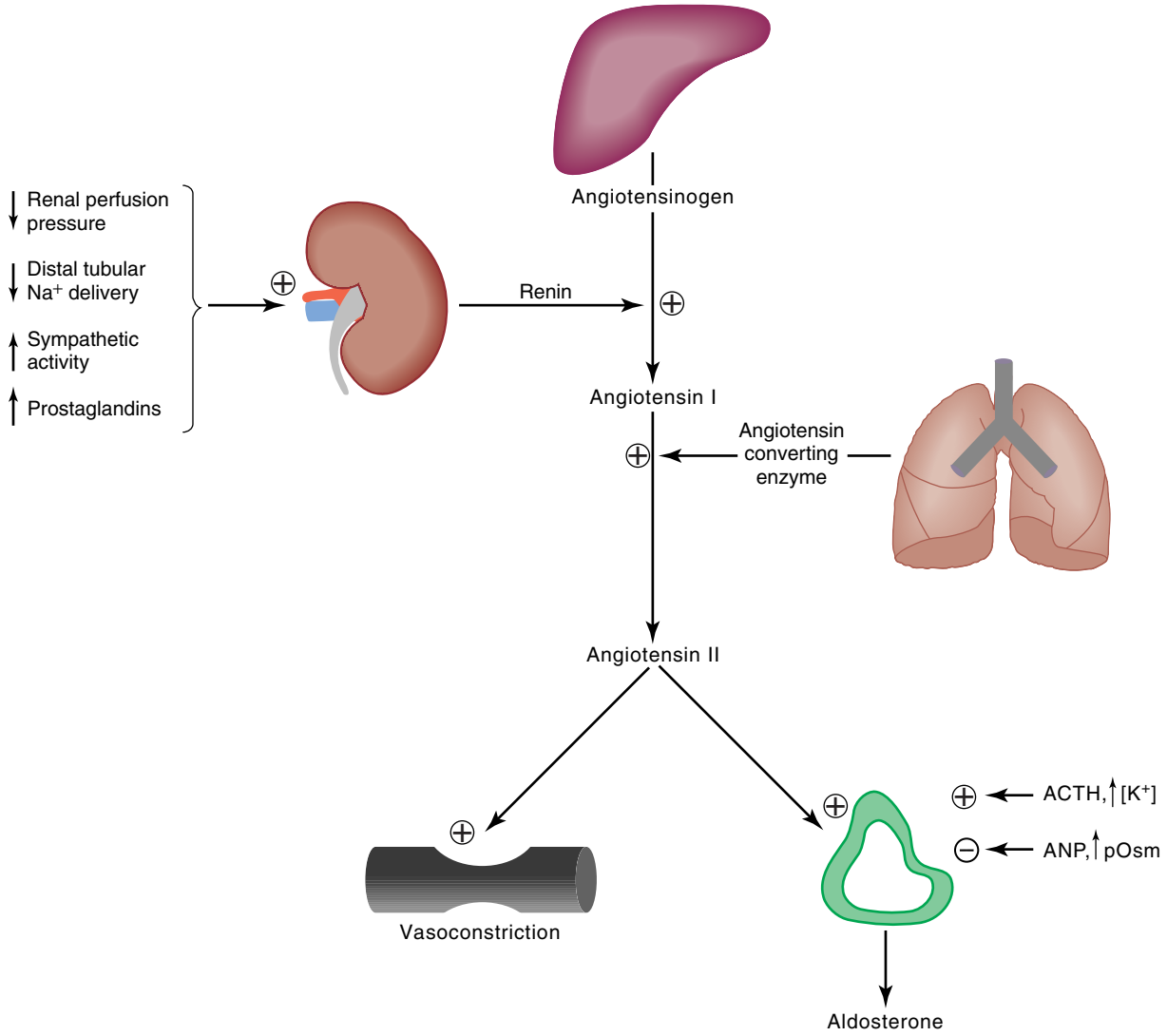


FIGURE 37.6 The renin–angiotensin cascade. Baroreceptors in the aortic arch, carotid sinus, and renal afferent arterioles sense hypovolemia and then cause the kidneys to secrete the enzyme renin. Renin cleaves angiotensinogen, which is synthesized by the liver, to produce angiotensin I. The angiotensin-converting enzyme, primarily in the lungs, cleaves angiotensin I to produce angiotensin II (AII), a peptide made of eight amino acid residues. AII is a potent vasoconstrictor and one of several stimulants of aldosterone secretion. Stimulation of aldosterone secretion from the adrenal cortex, along with possible direct intrarenal effects of AII, promotes renal conservation of sodium ions, complementing the pressor effect of AII in stabilizing arterial pressure and volume. ACTH, adrenocorticotropic hormone; ANP, atrial natriuretic peptide; pOsm, plasma osmolality.

thereby diminishes urinary excretion independent of AVP. Therefore, when water is the only drinking fluid available, an appropriate response of colloid-treated rats would be to stop drinking and thereby limit the secondary problem of osmotic dilution. In experiments, colloid-treated rats actually do stop drinking water despite persistent hypovolemia, and the stimulus for the inhibition of thirst has been found to be osmotic dilution of body fluids (Stricker, 1969). A 4 to 7% decrease in osmolality is sufficient to stop drinking that has been motivated by a 30 to 40% loss of plasma volume. Thus, the osmoregulatory system appears dominant in the control of thirst. Comparable data demonstrate the same to be true of AVP secretion in colloid-treated rats: osmotic dilution inhibits AVP secretion even in severely hypovolemic rats (Stricker & Verbalis, 1986).

Hypovolemia Also Stimulates Aldosterone Secretion and Salt Appetite

As mentioned earlier, hypovolemic animals need to consume and retain water and NaCl, not just water. Appropriately, colloid-treated rats drink NaCl solution

and conserve Na^+ in urine. Renal Na^+ retention is mediated largely by aldosterone secreted from the adrenal cortex, although Na^+ conservation also occurs in association with the decrease in GFR. The central nervous system does not innervate the adrenal cortex, as it does the adrenal medulla. However, a neural influence on aldosterone secretion is provided indirectly because AII is a very potent stimulus of aldosterone secretion, and renin secretion from the kidneys during hypovolemia is stimulated in part by the sympathetic nervous system (Fig. 37.6). The secretion of aldosterone is also stimulated by another peptide hormone, *adrenocorticotrophic hormone* (ACTH), which is secreted from the anterior lobe of the pituitary gland in response to *corticotropin-releasing hormone* (CRH). The release of CRH from the PVN is triggered by various stressors, including activated cardiovascular baroreceptors. Yet another stimulus of aldosterone secretion is increased plasma levels of K^+, which can develop as a consequence of reduced GFR and an associated decrease in urinary K^+ excretion. The effects of these three independent stimuli of aldosterone secretion are additive. Aldosterone has also been found to stimulate salt appetite in animals, which therefore complements its renal effects to cause sodium conservation. (See Box 37.4 on salt appetite after adrenocortical dysfunction.)

The onset of salt appetite induced in rats by colloid treatment is curiously delayed relative to the appearance of thirst. That is, water intake increases within 1 to 2 h after colloid treatment whereas the intake of NaCl solution does not increase until at least 5 h later (Stricker, 1981). Investigations showed that the delay is caused both by the gradual appearance of an excitatory stimulus of salt appetite (i.e., AII) and by the gradual disappearance of an inhibitory stimulus for NaCl intake (i.e., central secretion of oxytocin). When rats are maintained on Na^+-deficient diet for a few days instead of standard Na^+-rich laboratory chow, NaCl intake appears very rapidly after colloid treatment—even before thirst. Those results have been attributed in part to the suppression of central oxytocin secretion and in part to the increased blood levels of AII and aldosterone observed then; indeed, the latter two hormones are known to act synergistically in the stimulation of salt appetite.

Aldosterone and other mineralocorticoids appear to act directly on neurons in the NTS with mineralocorticoid receptors, which have the important additional property of containing an enzyme 11β-hydroxysteroid dehydrogenase type 2 (HSD2) that inactivates glucocorticoids (which also bind to those receptors) and thereby allows the receptors specificity to mineralocorticoids. These HSD2-containing cells are activated by all treatments that elicit salt appetite in rats but are quickly inactivated once saline has been consumed (Geerling, Engeland, Kawata, & Loewy, 2006). In addition, it seems relevant that rats with discrete lesions of the area postrema, which sends inhibitory neural input to those cells in the NTS, consumed prodigious amounts of concentrated NaCl solution whereas rats with area postrema lesions that invaded the NTS had much smaller saline intakes.

Analogous to the control of thirst, presystemic signals appear to inhibit salt appetite in association with

BOX 37.4

SALT APPETITE AFTER ADRENOCORTICAL DYSFUNCTION

Fifty years ago, a 4-year-old boy was brought to the Johns Hopkins University Hospital in Baltimore with the peculiar behavior of eating large amounts of table salt. The child heavily salted all his food, even juice, and often consumed salt directly from its container. His parents tried unsuccessfully to keep him from ingesting salt and brought him to the hospital in the hope that his strange behavior could be understood and stopped. The physicians were able to prevent the boy from having access to NaCl, but they were shocked when he died a few days later. An autopsy revealed that the child had bilateral tumors in his adrenal glands. In retrospect, it seems clear that he could not secrete aldosterone and therefore lost Na^+ in urine uncontrollably. Like patients with diabetes insipidus, who excrete a copious amount of dilute urine and drink comparable volumes of water in compensation (see Box 37.2 on diabetes insipidus), this boy evidently was consuming salt as an adaptive compensation to his recurrent need for sodium (Wilkins & Richter, 1940). He died when that response was prevented.

Rats increase their consumption of NaCl after surgical removal of their adrenal glands, by amounts proportional to the sodium loss in urine (Richter, 1936). The stimulus for this salt appetite appears to be a combination of elevated circulating levels of AII and reduced activity in a central oxytocinergic inhibitory system. In humans, however, a comparable salt appetite is not as common in patients with Addison disease, whose adrenocortical dysfunction causes them to lose Na^+ in urine excessively. Nor is salt appetite usually stimulated in people who lose excessive amounts of Na^+ in perspiration. No explanation is available for why an adaptive salt appetite appears readily in laboratory animals but not in humans, but this difference may be related, in part, to factors associated with herbivorous versus carnivorous eating behaviors.

gastrointestinal distension and increased plasma osmolality. The visceral signals appear to project through the vagus to the NTS and area postrema and then to the parabrachial nucleus (PBN), and, in this regard, lesions of peripheral sensory fibers or of the viscerosensory input to the PBN led to remarkably large intakes of NaCl solution by rats. In contrast, the inhibition of salt appetite associated with increased plasma osmolality has been attributed to central secretion of oxytocin.

Central Oxytocin Inhibits Salt Appetite

Recall that hypovolemic rats ordinarily drink water before they consume saline and, in doing so, thereby cause an osmotic dilution of body fluids. That dilution inhibits secretion of pituitary AVP and oxytocin, either of which effects might have removed an inhibitory stimulus for salt appetite. In other studies, rats showed a strong inverse relationship between intake of NaCl and plasma levels of oxytocin (but not AVP), suggesting that circulating oxytocin was an inhibitory stimulus of salt appetite. However, in tests of this hypothesis, an intravenous infusion of physiological doses of oxytocin did not decrease NaCl intake in hypovolemic rats, nor did infusion of an oxytocin receptor blocker increase NaCl intake. These findings were clarified by the observation that, coincident with the secretion of oxytocin from magnocellular neurons, oxytocin was released from parvicellular neurons projecting centrally from the PVN. Thus, plasma oxytocin may have been a peripheral marker of the centrally acting oxytocin that mediated the inhibition of salt appetite in rats (Blackburn, Verbalis, & Stricker, 1992). This hypothesis has been strongly supported by the results of a series of investigations. For example, salt appetite in hypovolemic rats was eliminated by a systemic injection of naloxone, an opioid receptor antagonist that disinhibits oxytocin secretion, and this effect was blocked by the prior injection of an oxytocin receptor antagonist directly into the cerebrospinal fluid. Conversely, NaCl ingestion in response to hypovolemia or AII was potentiated by diverse treatments that inhibit the secretion of oxytocin. In addition to osmotic dilution, these treatments included systemic injection of ethanol, maintenance on a sodium-deficient diet (instead of the standard laboratory diet rich in Na^+), and peripheral administration of mineralocorticoid hormones such as aldosterone. Thus, excitatory and inhibitory components together regulate NaCl intake in a dual-control system in the brain. See Box 37.5 on the control of thirst and salt appetite during hypovolemia in rats.

Central oxytocinergic neurons can inhibit salt intake and food intake (Chapter 36). Thus, when osmotic dehydration activates oxytocinergic neurons, the intake of food and salt is inhibited and hyperosmolality is thereby prevented. From this perspective, food is a source of osmolytes (not just of calories) and NaCl solution provides osmolytes without calories. Inhibition of intake of osmolytes complements the natriuretic effect of pituitary oxytocin in supporting osmoregulation.

Salt Appetite Is Also Inhibited by Atrial Natriuretic Peptide

The role of oxytocin in osmoregulation has become apparent only during the last decade or so, and the identification of other peptide hormones as important factors in the homeostasis of body fluids may be anticipated. One likely candidate is atrial natriuretic peptide. Like oxytocin, ANP is secreted by neurons within the brain, and when administered directly into the cerebrospinal fluid it inhibits an experimentally induced salt appetite in rats, perhaps by opposing the actions of AII. Moreover, destruction of ANP receptors in the brain eliminates the inhibition of salt appetite caused by an NaCl load (Blackburn, Samson, Fulton, Stricker, & Verbalis, 1995). Further work will be needed to understand the role of ANP-containing neurons in the central control of salt intake and the relationship of the neurons with other systems that participate in the regulation of water and NaCl intake.

Summary

Regulation of blood volume, like osmoregulation, is accomplished by a combination of physiological responses to hypovolemia, resulting in antidiuresis and antinatriuresis, and the complementary behavioral responses of increased water and NaCl intake. Cardiovascular baroreceptor cells detect hypovolemia and send neural signals to the brainstem, which communicates to the hypothalamus and forebrain structures mediating neurohypophyseal AVP and oxytocin secretion, as well as thirst. AII also appears to stimulate these responses while additionally stimulating salt appetite and supporting blood pressure as a potent vasoconstrictor. However, AII and hypovolemia both increase central oxytocin secretion, which mediates inhibition of NaCl intake, so for salt appetite to emerge this effect must be inhibited (as by osmotic dilution of body fluids, resulting from the renal retention of ingested water). The integration of these stimuli ensures that behavioral and physiological responses occur simultaneously, and their redundancy allows these vital regulatory processes to occur even when one stimulus is lost due to injury or disease. More generally, studies of water and NaCl ingestion provide insights into how the brain controls motivation, and how peptide and steroid hormones interact with neural signals in the control of behavior.

BOX 37.5

THIRST AND SALT APPETITE DURING HYPOVOLEMIA IN RATS

Hypovolemia elicits two endocrine effects in the excitation of salt appetite in rats—acute stimulation by a peptide hormone (AII) and more gradual stimulation by a steroid hormone (aldosterone). However, hypovolemia and AII each provide conflicting stimuli for salt appetite, since each signal also stimulates central oxytocin secretion, which mediates inhibition of salt appetite. The coordination of thirst and salt appetite stimulated by hypovolemia-induced colloid treatment in rats can be conceptualized as follows (Fig. 37.7). The combination of hypovolemia and AII stimulates thirst but provides a mixed stimulus of salt appetite. Thus, the animals at first drink water; however, by doing so, they dilute their body fluids, and the osmotic dilution eventually becomes large enough to inhibit thirst. Osmotic dilution has the additional effect of reducing central oxytocin secretion, which in turn disinhibits salt appetite, and the rats begin to drink concentrated NaCl solution. The ingested solute raises plasma osmolality and thereby removes the dilution-induced inhibition of thirst and oxytocin secretion. Consequently, the rats stop drinking saline and resume drinking water. Osmotic dilution again develops, thirst is again inhibited, and salt appetite is again disinhibited. And thus the hypovolemic animals, stimulated by neural and endocrine signals of hypovolemia, alternate their intakes of the two fluids while maintaining the concentration of body fluids near isotonic. Water and Na^+ are conserved in urine until the deficit in plasma volume is repaired, at which point the stimuli for adaptive physiological and behavioral responses disappear, and normal body fluid volume and tonicity are restored. Because the same neural and endocrine signals of hypovolemia stimulate thirst, salt appetite, and AVP and oxytocin secretion, these responses are integrated.

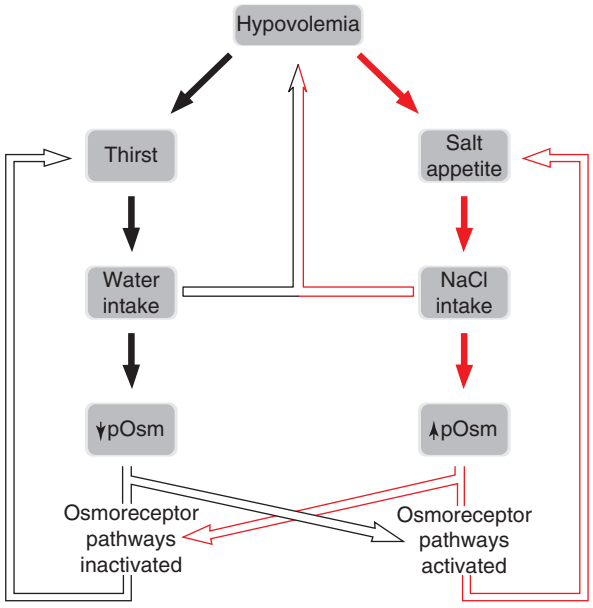


FIGURE 37.7 Schematic diagram of the mechanisms controlling thirst and salt appetite in hypovolemic rats. Solid arrows indicate stimulation and unfilled arrows indicate inhibition. The combination of the effects of blood-borne AII on the brain and neural baroreceptor signals to the brainstem stimulates hypovolemic animals to drink water and concentrated saline solution. The rats alternately drink the two fluids in amounts that ultimately add up to a volume of isotonic saline sufficient to repair the volume deficit. When the animals have access to only one fluid, the water and salt intakes are limited by activation of the appropriate inhibitory osmoregulatory pathways. When rats drink isotonic saline instead of water and concentrated saline, neither inhibitory pathway is activated and consequently intake is continuous.

In addition, presystemic signals may participate in the control of water and saline consumption. For example, distension of the stomach and small intestine by ingested fluid may inhibit further drinking despite continued deficits of plasma volume. Similarly, osmotic dilution detected by visceral osmoreceptors may inhibit intake of water, whereas osmotic concentration may inhibit intake of hypertonic NaCl solution. Thus, both systemic and presystemic signals, providing information about the concentration and volume of body fluid and of ingested fluid, respectively, appear to be available to the hypovolemic animal. The integration of these signals by the brain is critical in the control of water and saline intakes and determines the volume of each fluid that is consumed.

References

Blackburn, R. E., Samson, W. K., Fulton, R. J., Stricker, E. M., & Verbalis, J. G. (1995). Central oxytocin and atrial natriuretic peptide receptors mediate osmotic inhibition of salt appetite in rats. *American Journal of Physiology, 269,* R245–R251.

Blackburn, R. E., Verbalis, J. G., & Stricker, E. M. (1992). Central oxytocin mediates inhibition of sodium appetite by naloxone in hypovolemic rats. *Neuroendocrinology, 56,* 255–263.

Egan, G., Silk, T., Zamarripa, F., Williams, J., Federico, P., Cunnington, R., … , Denton, D. (2003). Neural correlates of the emergence of consciousness of thirst. *Proceedings of the National Academy of Sciences of the United States of America, 100,* 15241–15246.

Ferguson, A. V., & Renaud, L. P. (1986). Systemic angiotensin acts at subfornical organ to facilitate activity of neurohypophysial neurons. *American Journal of Physiology, 251,* R712–R717.

Fitzsimons, J. T. (1961). Drinking by rats depleted of body fluid without increase in osmotic pressure. *Journal of Physiology (London), 159,* 297–309.

Fitzsimons, J. T. (1969). The role of a renal thirst factor in drinking induced by extracellular stimuli. *Journal of Physiology (London), 155,* 563–579.

Fraser, C. L., & Arieff, A. I. (1997). Epidemiology, pathophysiology, and management of hyponatremic encephalopathy. *American Journal of Medicine, 102,* 67–77.

Geerling, J. C., Engeland, W. C., Kawata, M., & Loewy, A. D. (2006). Aldosterone target neurons in the nucleus tractus solitarius drive sodium appetite. *Journal of Neuroscience, 26,* 411–417.

Gilman, A. (1937). The relation between blood osmotic pressure, fluid distribution and voluntary water intake. *American Journal of Physiology, 120,* 323–328.

Hoffman, G. E., Smith, M. S., & Verbalis, J. G. (1993). C-Fos and related immediate early gene products as markers of activity in neuroendocrine systems. *Frontiers in Neuroendocrinology, 14,* 173–213.

Johnson, A. K., & Buggy, J. (1978). Periventricular preoptic-hypothalamus is vital for thirst and normal water economy. *American Journal of Physiology, 234,* R122–R129.

Kanter, G. S. (1953). Excretion and drinking after salt loading in dogs. *American Journal of Physiology, 174,* 87–94.

Knepper, M. A. (1997). Molecular physiology of urinary concentrating mechanism: Regulation of aquaporin water channels by vasopressin. *American Journal of Physiology, 272,* F3–F12.

Noakes, T. D., Sharwood, K., Speedy, D., Hew, T., Reid, S., Dugas, J., … , Weschler, L. (2005). Three independent biological mechanisms cause exercise-associated hyponatremia: evidence from 2,135 weighed competitive athletic performances. *Proceedings of National Academy of Science of United States of America, 102,* 18550–18555.

Richter, C. P. (1936). Increased salt appetite in adrenalectomized rats. *American Journal of Physiology, 115,* 155–161.

Robertson, G. L. (1986). Posterior pituitary. In P. Felig, J. Baxter & L. A Frohman (Eds.), *Endocrinology and metabolism* (pp. 338–385). New York, NY: McGraw-Hill.

Robinson, A. G. (1985). Disorders of antidiuretic hormone secretion. *Clinical Endocrinology Metabolism, 14,* 55–88.

Robinson, M. M., & Evered, M. D. (1987). Pressor action of intravenous angiotensin II reduces drinking response in rats. *American Journal of Physiology, 252,* R754–R759.

Starling, E. H. (1896). On the absorption of fluids from the connective tissue spaces. *Journal of Physiology (London), 19,* 312–326.

Stricker, E. M. (1969). Osmoregulation and volume regulation in rats: Inhibition of hypovolemic thirst by water. *American Journal of Physiology, 217,* 98–105.

Stricker, E. M. (1976). Drinking by rats after lateral hypothalamic lesions: A new look at the lateral hypothalamic syndrome. *Journal of Comparative Physiological Psychology, 90,* 127–143.

Stricker, E. M. (1981). Thirst and sodium appetite after colloid treatment in rats. *Journal of Comparative Physiological Psychology, 95,* 1–25.

Stricker, E. M., & Verbalis, J. G. (1986). Interaction of osmotic and volume stimuli in regulation of neurohypophyseal secretion in rats. *American Journal of Physiology, 250,* R267–R275.

Thrasher, T. N., Keil, L. C., & Ramsay, D. J. (1982). Lesions of the laminar terminalis (OVLT) attenuate osmotically induced drinking and vasopressin secretion in the dog. *Endocrinology (Baltimore), 110,* 1837–1839.

Thrasher, T. N., Nistal-Herrera, J. F., Keil, L. C., & Ramsay, D. J. (1981). Satiety and inhibition of vasopressin secretion in dogs. *American Journal of Physiology, 240,* E394–E401.

Verbalis, J. G. (2007). The syndrome of inappropriate antidiuretic hormone secretion and other hypoosmolar disorders. In R. W. Schrier (Ed.), *Diseases of the kidney and urinary tract* (pp. 2214–2248). Philadelphia: Lippincott Williams & Wilkins.

Verbalis, J. G. (2010). Brain volume regulation in response to changes in osmolality. *Neuroscience, 168,* 862–870.

Verbalis, J. G., Mangione, M. P., & Stricker, E. M. (1991). Oxytocin produces natriuresis in rats at physiological plasma concentrations. *Endocrinology (Baltimore), 128,* 1317–1322.

Verney, E. B. (1947). The antidiuretic hormone and the factors which determine its release. *Proceedings of the Society (London) B135*, 25–105.

Wilkins, L., & Richter, C. P. (1940). A great craving for salt by a child with cortico-adrenal insufficiency. *Journal of American Medical Association, 114,* 866–868.

Suggested Readings

Denton, D. (1982). *The hunger for salt: an anthropological, physiological and medical analysis*. Berlin: Springer-Verlag.

Fitzsimons, J. T. (1979). *The physiology of thirst and sodium appetite*. Cambridge: Cambridge University Press.

Stricker, E. M. (2004). Thirst. In E. M. Stricker & S. C. Woods (Eds.), *Handbook of behavioral neurobiology* (Vol. 14, pp. 505–543). New York, NY: Kluwer Academic/Plenum.

Verbalis, J. G. (1990). Clinical aspects of body fluid homeostasis in humans. In E. M. Stricker (Ed.), *Handbook of behavioral neurobiology* (Vol. 10, pp. 421–462). New York, NY: Plenum.

Wolf, A. V. (1958). *Thirst: physiology of the urge to drink and problems of water lack*. Springfield, IL: Thomas.

Edward M. Stricker and Joseph G. Verbalis

This page intentionally left blank

CHAPTER

38

Neuroendocrine Systems

THE HYPOTHALAMUS IS THE CENTRAL NEUROENDOCRINE ORGAN

The hypothalamus, located at the most rostral region of the brainstem in the diencephalon, is a key center regulating numerous and diverse physiological functions, including growth, metabolism, stress responses, reproduction, osmoregulation, and circadian rhythms. All these functions are critically involved in maintaining the homeostasis of the organism and in coordinating the occurrence and timing of physiological functions with the appropriate environmental conditions. In order to accomplish this, the hypothalamus acts as an integrator, receiving converging inputs from the sensory and autonomic systems related to the internal and external environment of the organism. The hypothalamus responds rapidly to this convergent information by changing its output of neurotransmitters and neuropeptides, which in turn effect change at its targets within the central nervous system (CNS), the pituitary gland, and the bloodstream.

The hypothalamus contains several populations of cells that are defined as neuroendocrine, meaning that they have both neuronal and endocrine features. The neuroendocrine cells express neural markers and release a peptide or neurotransmitter in response to a depolarizing stimulus, similar to other neurons in the brain. Unlike other "traditional" neurons, neuroendocrine cells release their neurotransmitter directly into a blood system. Considering that endocrine glands are defined by the release of a chemical into the bloodstream, neuroendocrine cells are part of the endocrine system. Although the chemicals released by most neuroendocrine cells are peptides and neurotransmitters, they are also appropriately referred to as hormones, or neurohormones, because of this endocrine property.

The primary target of hypothalamic neuroendocrine cells is the pituitary gland, comprising the anterior pituitary (adenohypophysis) and the posterior pituitary (neurohypophysis). Some species also possess an intermediate pituitary lobe. The hypothalamic cells that regulate the anterior pituitary are referred to as hypothalamic releasing or inhibiting cells, and are the major focus of this chapter. These neurons have their cell somata in the hypothalamus, and they have an axonal process that extends to and terminates in the median eminence at the base of the hypothalamus (Yin & Gore, 2010). The median eminence is highly vascularized by a blood system called the portal capillary plexus. Hypothalamic releasing/inhibiting hormones are released into the portal capillary vasculature, which then transports these neurohormones to the anterior pituitary gland (Fig. 38.1). There, each hypothalamic hormone has at least one corresponding pituitary hormone whose release is regulated by patterns and levels of hypothalamic input.

The posterior pituitary gland also is regulated by hypothalamic cells, but the anatomical connectivity between the hypothalamus and the neurohypophysis differs from the hypothalamic-adenohypophysial system (Fig. 38.1). Hypothalamic cells that regulate the neurohypophysis have their cell bodies in the hypothalamus, but their nerve terminals project directly into the posterior pituitary gland. From there, the neurohypophyseal hormones are released into the general circulatory system and are transported to reach their targets elsewhere in the body. This chapter begins with a detailed description of hypothalamic-anterior pituitary control, followed by a section on hypothalamic-posterior pituitary control (Box 38.1).

© 2013 Elsevier Inc. All rights reserved.

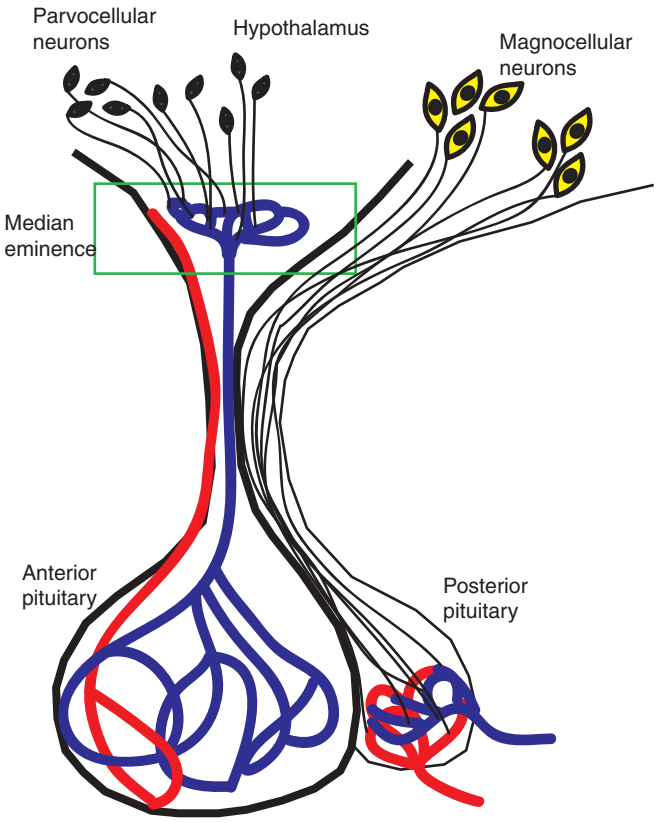


FIGURE 38.1 Overview of the central neuroendocrine systems. Several populations of hypothalamic cells regulate anterior pituitary function through release of releasing or inhibiting factors into the portal capillary vasculature. Cell bodies of these hypothalamic-adenohypophysial are small in size (parvocellular), and terminate in the median eminence at the base of the hypothalamus. Other populations of hypothalamic neurons are much larger (magnocellular), existing in the paraventricular and supraoptic nuclei, and project into the posterior pituitary (hypophysis) to release their hormones into general circulation.

HYPOTHALAMIC CONTROL OF ANTERIOR PITUITARY HORMONES AND THEIR REGULATED FUNCTIONS

The hypothalamic regulation of anterior pituitary hormonal synthesis and release is integral to the regulation of stress, reproduction, metabolism, growth, and lactation (Table 38.1). Each of these systems has a similar anatomical and hierarchical organization (Fig. 38.1). At the top of the hierarchy are the hypothalamic neurons that release their neurohormone from neuroterminals in the median eminence. With the exception of the prolactin-inhibiting hormone (dopamine), all of the hypothalamic releasing/inhibiting hormones are peptides that are cleaved from larger prohormones. The processed hypothalamic hormones are released from secretory vesicles in the axon terminals and diffuse through the extracellular fluid into the portal capillary vasculature, by a mechanism similar to volume transmission. Once in the portal capillaries, the hormones are transported to the pituitary gland where they exert an effect on synthesis and secretion of one or more pituitary hormones through binding to a specific receptor on a subset of pituitary cells. The pituitary hormones are all large peptides or small proteins. Pituitary hormone release into the general circulation is stimulated by hypothalamic releasing hormones, and each pituitary hormone targets a tertiary organ to cause hormone release or a functional outcome. In the case of hypothalamic inhibiting hormones, pituitary hormone synthesis/secretion is inhibited.

The hypothalamic–pituitary target organ axes are not an independent series of parallel regulatory systems. These axes have considerable crosstalk with one another

BOX 38.1

HISTORICAL PERSPECTIVE ON HYPOTHALAMIC HORMONES

The concept of a link between the hypothalamus and the pituitary gland was suggested and later demonstrated by Geoffrey Harris (1971), who proved the existence of a tiny series of blood vessels leading from the median eminence at the base of the hypothalamus to the anterior pituitary gland. Harris proposed that the hypothalamus controls the release of anterior pituitary hormones via this vascular connection. Subsequent studies have shown that the hypothalamic–anterior pituitary connection is critical for maintaining numerous physiological functions involved in metabolism, growth, reproduction, stress responses, and lactation. The isolation and purification of the hypothalamic releasing hormones in the late 1960s by the laboratories of Roger Guillemin, Andrew Schally, and others led to an enormous increase in knowledge of these molecules. Guillemin and Schally were awarded the Nobel Prize in Physiology and Medicine in 1977 for these fundamental discoveries in neuroendocrinology. They shared it with Rosalyn Yalow, who invented the radioimmunoassay, a tool for carefully and accurately measuring hormone levels in the body.

Andrea C. Gore

TABLE 38.1 Summary of the Five Neuroendocrine Hypothalamic Systems Regulating Anterior Pituitary Function

Function	Hypothalamic Hormone(s)	Receptor Characteristic & Localization	Anterior Pituitary Hormone(s)	Receptor Characteristic & Localization	Target Hormone(s)	Receptor Characteristic & Localization
Stress	Corticotropin-releasing hormone (CRH): 41 amino acid peptide, cleaved from precursor	CRH receptor: GPCR; on corticotrope (also CNS, adrenal, placenta, ovary, immune tissues)	Adrenocorticotropic hormone (ACTH), 39 amino acid peptide, cleaved from precursor and encoded by POMC gene	ACTH receptor: melanocortin receptor family of GPCRs; in adrenal cortex	Cortisol (humans) or corticosterone (rats, mice); steroid hormone family	Glucocorticoid receptor; steroid receptor family member; in brain & numerous body tissues
Basal metabolism (thyroid)	Thyrotropin-releasing hormone (TRH): 3 amino acid peptide, multiple copies cleaved from precursor	TRH receptor: GPCR; on thyrotrope	Thyroid-stimulating hormone (TSH), composed of α and β subunits of ~100 amino acids each	TSH receptor: GPCR, mainly in thyroid gland	Thyroid hormones (T3 and T4); lipophilic hormones synthesized from tyrosine and iodide	Thyroid hormone receptor; nuclear receptor superfamily member; in CNS and most tissues
Growth	Growth-hormone releasing hormone (GHRH): 44 amino acid peptide, cleaved from precursor Growth-hormone inhibiting hormone (GHIH, also called somatostatin): 14 or 28 amino acid peptide, cleaved from precursor	GHRH receptor: GPCR; on somatotrope Somatostatin receptor: 5 known GPCRs; on somatotrope & other tissues	Growth hormone (GH, also called somatotropin), 191 amino acid protein	GH receptor: cytokine receptor family member, on liver, muscle, bone, & other tissues	Insulin-like growth factor-1 (IGF-1): 70 amino acid peptide, cleaved from precursor	IGF-1 receptor: receptor tyrosine kinase family member; located on muscle, bone, brain, & other tissues
Reproduction	Gonadotropin-releasing hormone (GnRH)*: 10 amino acid peptide, cleaved from precursor	GnRH receptor: GPCR; on gonadotropes, brain, gonads, other tissues	Luteinizing hormone (LH), Follicle-stimulating hormone (FSH): they have a common α and a unique β subunit of ~100 amino acids each	LH receptor, FSH receptor: GPCRs, on gonads	Gonadal steroids (estrogens, progestins, androgens)	Gonadal steroid receptors (ER, PR, AR): member of steroid receptor family; in reproductive tissues, brain, pituitary, & body tissues
Lactation	Dopamine (prolactin-inhibiting hormone)	Dopamine D2 receptor: GPCR, on lactotropes & CNS	Prolactin (PRL): 198 amino acid protein	PRL receptor: GH receptor cytokine receptor family member; on mammary gland (also liver, kidney, GI tract, brain, adrenal, gonad)	NA (milk is not a hormone within the lactating woman)	NA

*A gonadotropin-inhibiting hormone (GnIH) has recently been identified (Kriegsfeld et al., 2006), but it appears to act as a traditional neuropeptide rather than a neuroendocrine substance in mammals.

Abbreviations: AR, androgen receptor; CNS, central nervous system; ER, estrogen receptor; GPCR, G-protein coupled receptor; GI: gastrointestinal; NA: not applicable; POMC, proopiomelanocortin; PR, progesterone receptor.

and can interact at multiple levels. For example, reproductive and stress axes regulate each other at specific hypothalamic neurons, both through direct actions at the CNS, as well as through feedback from the target steroid hormones. During stressful situations, elevated adrenal glucocorticoids will feed back to the brain and have actions not only at the CRH neurons, but at the GnRH neurons as well, thereby suppressing reproduction at a time when it is not optimal. Therefore, these hypothalamic neurons serve as integrators of multiple, complex pieces of information about the homeostasis and environment of the organism.

The Hypothalamus is Regulated by Neural, Glial, and Hormonal Inputs

As already stated, hypothalamic neuroendocrine cells are at the top of the hierarchy in the control of homeostatic function. Nevertheless, it is important to remember that the hypothalamus is an integrated part of the central nervous system. Thus, the hypothalamus has neural afferents/efferents between hypothalamic nuclei, as well as with brain nuclei outside of the hypothalamus. For example, each hypothalamic-releasing or -inhibiting hormone neuron receives afferent inputs from neurons and is in close communication with glia. The importance of such neural/glia communication within and between hypothalamic populations, and between hypothalamic and other CNS regions, is made clear when considering the regulation of homeostasis. In order for the organism to respond to its environment, both internal and external, it must receive input from the environment and be able to integrate that input in a coordinated manner. This is certainly the case with hypothalamic neuroendocrine cells, which are responsive to scores of neurotransmitters and neurotrophic factors arising from numerous neural/glial sources.

Hypothalamic neuroendocrine cells are also regulated by feedback from the hormones released by peripheral target organs (Table 38.1). This may occur directly at the respective hypothalamic cells via the target hormone receptor. For example, stress-regulating CRH neurons express the glucocorticoid receptor, which can sense circulating adrenal cortisol levels. In addition or alternatively, some hypothalamic neurons may be regulated by indirect feedback from peripheral hormones, mediated by other interneurons. This appears to be the case for reproduction-regulating GnRH neurons, which do not express the major estrogen, progestin, or androgen receptors. For most neuroendocrine systems, feedback regulation is a combination of direct and indirect. In addition, feedback is generally negative, meaning that high levels of a peripheral target hormone downregulate the hypothalamic cells at the top of the hierarchy. When peripheral hormone levels are low, negative feedback is relieved or reduced, enabling the hypothalamic hormone to increase its output and stimulate that neuroendocrine axis to replenish/increase peripheral hormone concentrations. In the case of reproduction, there is also positive feedback of peripheral hormones in females during ovarian cycles, something that will be discussed in the hypothalamic-pituitary-gonadal section following.

Characteristics of Neurohormone Release: Ultradian and Circadian/Diurnal Patterns

In considering the release of a hypothalamic hormone and its resulting effect on downstream targets, it is important to note the rhythmicity of this release. Hypothalamic hormones are released in a pulsatile manner, characterized by brief bursts of release, usually at intervals of 1–2 h (Krsmanovic, Hu, Leung, Feng, & Catt, 2009). This rhythmic pattern also is referred to as ultradian (i.e., shorter than daily rhythms) or circhoral (approximately hourly intervals of release). The corresponding anterior pituitary cells respond to each pulse of a hypothalamic hormone with a corresponding pulse of its pituitary hormone shortly thereafter (Fig. 38.2). It has been speculated that the pulsatile release of hypothalamic hormones is necessary to prevent the desensitization of their receptors in the anterior pituitary gland.

Along with these ultradian pulses of hypothalamic pulses are the circadian or diurnal rhythms of hormonal

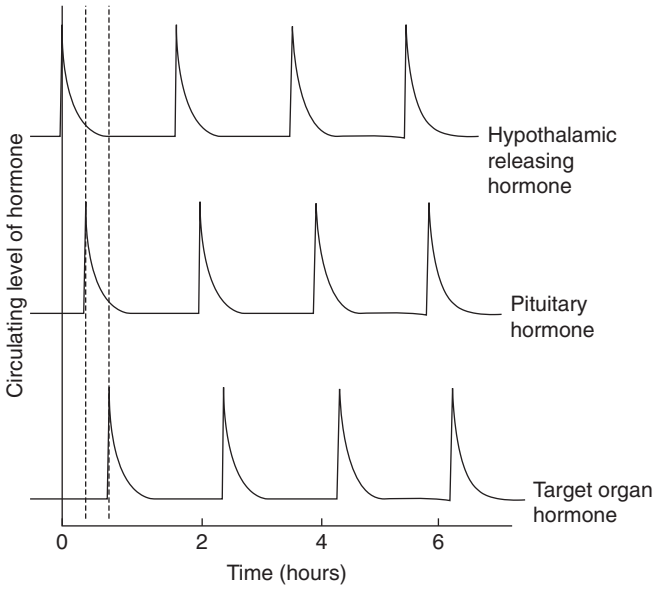


FIGURE 38.2 Representation of pulsatile hormone release in a neuroendocrine system. A representative hypothalamic-releasing hormone is released in discrete pulses at 1- to 2-h intervals. This causes pulses of the corresponding anterior pituitary hormone to occur shortly thereafter. The release of the pituitary hormone stimulates the target organ to release its hormones.

release that are characteristic of the neuroendocrine system. These 24-h rhythms of hormone release, which can occur independently of the light/dark cycle (circadian) or are entrained to the light/dark cycle (diurnal), are characteristic of neuroendocrine systems (Gore, 1998). Circadian rhythms are driven by a part of the hypothalamus called the suprachiasmatic nucleus (SCN), and the SCN makes projections within the hypothalamus to regulate neuroendocrine release in a 24-h cycle.

Cycles of hormonal release longer than 24 h can be observed for the reproductive axis. Female vertebrates often exhibit reproductive cycles, such as the four-day estrous cycles of rats or the 28-day menstrual cycles of many primates, including humans. Reproductive hormones undergo substantial fluctuations across these cycles. Even longer cycles are observed in the cases of seasonal breeders, as discussed following in the section on the hypothalamic control of reproduction.

Hypothalamic Control of Stress

Overview, Characteristics and Physiological Functions of the Hypothalamic-Pituitary-Adrenal Axis

In order to survive, an animal must be able to respond to stressors, an internal or external cue that disrupts the homeostatic status of the animal (Aguilera, 2011). The stress response enables the "fight-or-flight" reaction and is highly adaptive in the short term, although it may become maladaptive in the long term.

The neuroendocrine system involved in mediating the stress response is the hypothalamic–pituitary–adrenal (HPA) axis. As summarized in Table 38.1, the hypothalamic releasing hormone controlling this axis is corticotropin-releasing hormone (CRH), produced in highest levels in the medial parvocellular part of the paraventricular nucleus (PVN; Fig. 38.3). CRH nerve terminals release this 41–amino acid peptide into the portal vasculature, where CRH targets the corticotropes of the anterior pituitary gland. Additionally, arginine vasopressin (AVP), another hypothalamic hormone produced in the magnocellular elements of the PVN, acts at pituitary corticotropes to exert synergistic effects. The corticotropes have receptors for and respond to both CRH and AVP with the synthesis and release of adrenocorticotropic hormone (ACTH). ACTH is released from the pituitary gland into the general circulation, where it acts at receptors on the adrenal cortex to stimulate the production and release of glucocorticoids. The feedback of glucocorticoids onto the brain and pituitary negatively regulates the synthesis of CRH and ACTH: therefore, this neuroendocrine system is a classical negative feedback loop.

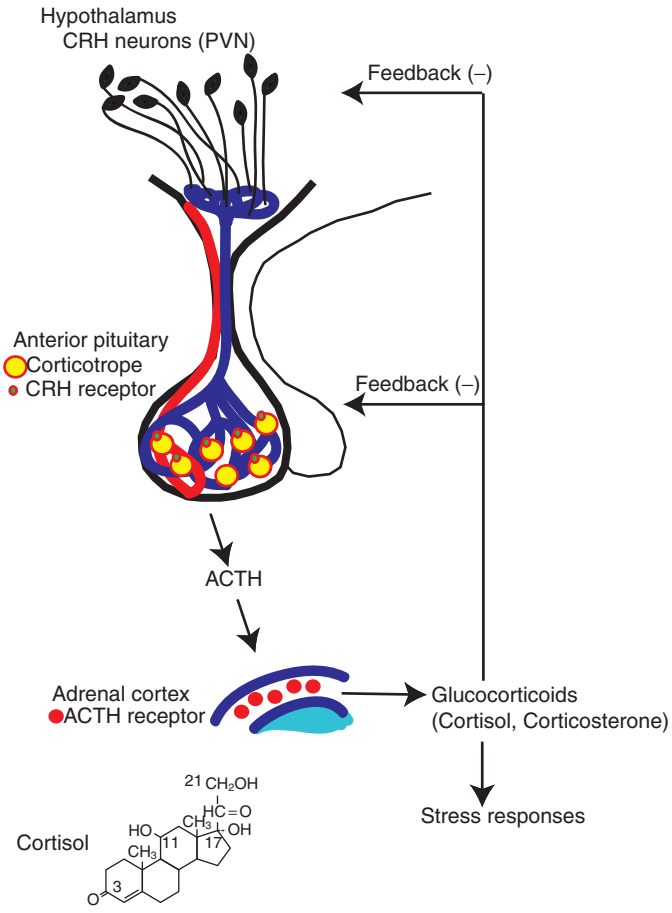


FIGURE 38.3 The hypothalamic–pituitary–adrenal (HPA) stress axis is shown. Hypothalamic CRH neurons in the paraventricular nucleus project their neuroterminals to the external zone of the median eminence. CRH travels through the portal vasculature to cause the synthesis and release of ACTH from corticotropes. This molecule acts at the adrenal cortex to cause the biosynthesis and release of glucocorticoids. These are released into general circulation to mediate the stress response, and they feed back in a negative manner at the corticotropes of the pituitary and the CRH neurons of the PVN (indicated by long arrows). The structure of cortisol, the major glucocorticoid released by the adrenal cortex in humans, is shown (bottom). Abbreviations: CRH: corticotropic-releasing hormone; ACTH: adrenocorticotropic hormone; PVN: paraventricular nucleus.

PHYSIOLOGY

Corticosteroids are necessary for normal fetal development. Enzymes in the liver, gastrointestinal system, lungs, and adrenal medulla are stimulated by glucocorticoids. There, glucocorticoids play important roles in the regulation of glycogen utilization and storage, with glucocorticoids promoting the production of glycogen synthase. The stress axis is subject to developmental and age-related regulation due to changes in afferent connections to the hypothalamus, glucocorticoid and mineralocorticoid receptor-binding changes, and alterations in the negative feedback of glucocorticoids on CRH neurons.

Glucocorticoids exert effects on the cardiovascular system, causing elevations in heart rate and blood pressure and alterations in blood flow. They act in close coordination with the autonomic nervous system to exert their effects. In general, their role is to mobilize energy stores and to improve cardiovascular tone. Glucocorticoids may also modulate immune responses and inhibit cytokine production, thereby playing a role in the inhibition of the inflammatory response. Glucocorticoids can interact with hormones produced by other neuroendocrine axes, such as those involved in thyroid function, reproductive function, and growth (O'Connor, O'Halloran, & Shanahan, 2000).

RELEASE

Stress hormones exhibit both pulsatile and circadian cycles. When the hypothalamic nucleus responsible for circadian rhythms, the SCN, is lesioned, the 24-h pattern of CRH release is obliterated. During the 24-h day, glucocorticoid levels peak during the early morning (~0700 h) and reach a nadir during the afternoon in humans.

Glucocorticoids and their Receptors

Glucocorticoids are members of the steroid hormone family. As is the case for sex steroid hormones (discussed later), they derive from a cholesterol precursor. Following a series of enzymatic reactions, the major corticosteroids, such as cortisol (in humans) or corticosterone (in rats and mice), are synthesized.

The glucocorticoid receptor (GR) is widely distributed in the central nervous system, including the hypothalamus and, more specifically, the CRH cells of the PVN, where cortisol negative feedback can be exerted. GRs are also abundant in the hippocampus, in pyramidal and dentate granule cells. The hippocampus appears to play a role in mediating the stress response, and there are multiple redundant pathways between the hippocampus and the PVN. GRs are found in many other brain regions such as the olfactory cortex, cerebral cortex, amygdala, septum, thalamus, midbrain raphe nuclei, cerebellum, and brainstem (Morimoto, Morita, Ozawa, Yokoyama, & Kawata, 1996). The widespread distribution of this receptor suggests important and diverse effects of glucocorticoids in the central nervous system.

Glucocorticoids can also bind to the mineralocorticoid receptor (MR), the primary ligand of which is aldosterone. Moreover, the brain MR has a higher (~10-fold higher) affinity for glucocorticoids than the GR. Both of these receptors are members of the steroid hormone superfamily. They are thought to reside in the cytoplasm of target cells in the unliganded state. When bound to their ligands, GRs or MRs translocate into the nucleus, where they bind to specific sites on the promoter of the responsive gene. Thus, GR and MR are transcription factors. Their efficacy in stimulating or inhibiting gene transcription is modulated by their binding to other molecules, such as heat shock proteins, which can alter the translocation of the receptor–ligand complex to the nucleus, or transcriptional coactivators, which modulate the complexes associated with the gene.

Glucocorticoids are involved in numerous neural processes, both normal and pathological. Chronic high glucocorticoid exposure may lead to neurodegeneration and eventually cell death. A lack of glucocorticoids can also be detrimental. In the dentate gyrus, maintenance of the dentate granule cell number appears to require glucocorticoids, and adrenalectomy results in the loss of granule cells. Moreover, neurogenesis in the adult dentate gyrus also appears to be glucocorticoid dependent. There are functional correlates to the effects of glucocorticoids on the hippocampus, with these hormones playing roles in learning and memory. Acute stress can interfere with the ability of an organism to learn, and chronic stress is correlated with deficits in spatial learning.

Disruptions in the Stress Axis

The negative feedback loop of CRH–ACTH–glucocorticoids is kept in a delicate balance. In response to stress, there is a large increase in the activity of the stress axis, but the system is downregulated rapidly by negative feedback from the glucocorticoids to the brain and pituitary gland, causing it to return the output of the stress axis to basal low levels.

Overactivity of the stress axis can increase susceptibility to infection and tumors and reduce inflammatory and autoimmune disease (O'Connor et al., 2000). For example, Cushing's disease is characterized by hypercortisolemia and abnormal secretion and/or bioactivity of ACTH. The high levels of cortisol suppress CRH release and cause alterations in immune function. People with Cushing's disease have an increase in appetite and can experience changes in fat distribution.

Underactivity of the stress axis can cause the opposite effects of hyperactivity—namely resistance to infection and tumors but increased susceptibility to inflammatory and autoimmune disease. In the case of adrenocortical insufficiency, there is generally a failure of the adrenal cortex to respond to ACTH. Addison's disease is characterized by an insufficient release of glucocorticoid secretion, and its symptoms include muscle weakness and fatigue, psychological symptoms, gastrointestinal symptoms, and changes in skin pigmentation. Fortunately, glucocorticoid and mineralocorticoid replacement can help mitigate these symptoms.

Hypothalamic Control of Basal Metabolism

Overview, Characteristics, and Physiological Functions of the Hypothalamic-Pituitary-Thyroid Axis

The proper regulation of metabolism is critical for the development and survival of the organism. The

hypothalamic–pituitary–thyroid (HPT) axis plays a critical role in the control of metabolic activity, stimulating an increase in metabolic processes in target cells. The hypothalamic hormone involved in this function is thyrotropin-releasing hormone (TRH; Nillni, 2010), produced by a group of cells in a specific nucleus in the hypothalamus, the paraventricular nucleus, parvocellular compartment (Fig. 38.4). TRH targets receptors on thyrotropes, a subset of cells in the anterior pituitary gland that, when stimulated, produce the thyroid-stimulating hormone (TSH). TSH acts at receptors in the thyroid gland, located in the neck. These cells then produce the thyroid hormones, thyroxine (T_4) and triiodothyronine (T_3), the latter being the more potent thyroid hormone, which are released into the general circulation.

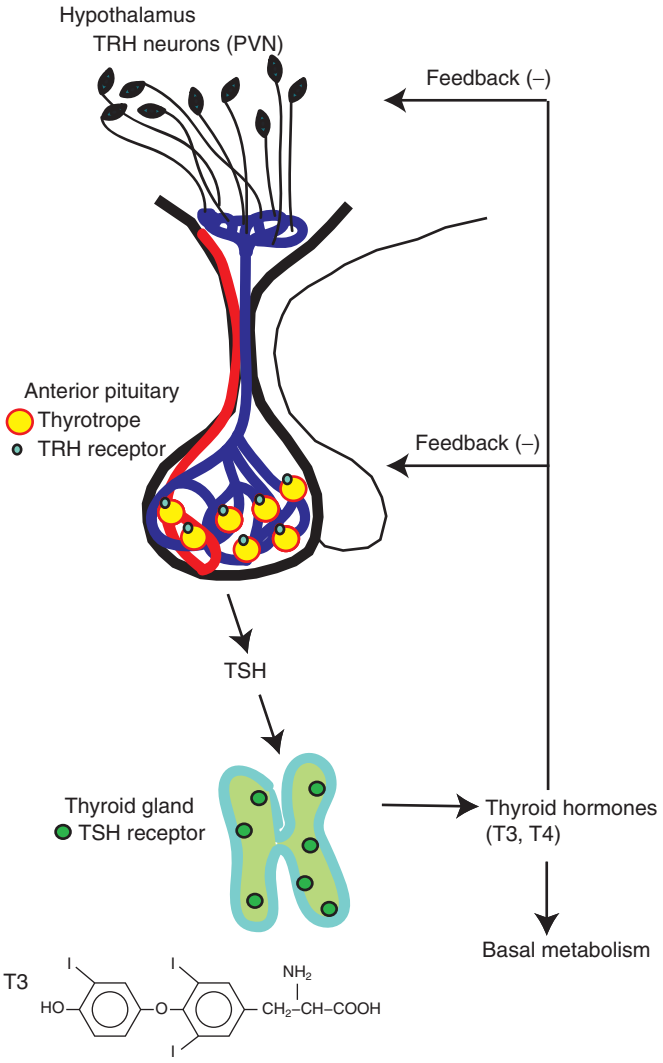


FIGURE 38.4 The hypothalamic–pituitary–thyroid (HPT) axis controlling basal metabolic functions is shown. TRH neurons in the paraventricular nucleus of the hypothalamus project their neuroterminals to the external zone of the median eminence. TRH travels through the portal vasculature to cause the synthesis and release of TSH from thyrotropes. This molecule acts at the thyroid gland to cause the biosynthesis and release of the thyroid hormones, T_3 and T_4. These are released into the general circulation to affect metabolism. Negative feedback of T_3 and T_4 at the thyrotropes of the pituitary and at the TRH neurons of the PVN is indicated by the long arrows. The structure of T_3 is shown (bottom). Abbreviations: TRH: thyrotropin-releasing hormone; TSH: thyroid-stimulating hormone; T_3: triiodothyronine; T_4: tetraiodothyronine or thyroxine; PVN: paraventricular nucleus.

PHYSIOLOGY

Thyroid hormones act at target organs throughout the body to regulate cellular metabolism. T_3 and T_4 also feed back at the brain and anterior pituitary gland to regulate levels of TRH and TSH, respectively. The thyroid hormone axis is a classical negative feedback loop, as thyroid hormone feedback inhibits the secretion of these hypothalamic and pituitary hormones, thereby maintaining appropriate levels of free thyroid hormone in the bloodstream. Such precise regulation of thyroid hormone levels is critically important because elevated or depressed levels of this hormone can have substantial adverse consequences on the body, particularly during sensitive periods of development.

The brain–pituitary–thyroid axis is involved in metabolism, protein synthesis, and development of the brain. Thyroid hormone receptors are located throughout the brain, and they are also found in peripheral and autonomic nervous systems, pituitary gland, bone, muscle, liver, heart, testis, lung, placenta, and other tissues. The major function of the thyroid hormone is the regulation of basal metabolic rate and calorigenic responses. In the target cells, thyroid hormones regulate ATP levels produced by mitochondria and alter protein synthesis. The maintenance of thyroid hormone levels is extremely important for adult metabolic activity, and thyroid hormone abnormalities in the developing organism can have catastrophic consequences. Thyroid hormone deprivation in the fetus causes permanent nervous system abnormalities and substantial adverse effects on growth and bone development that may be reversed or mitigated only if the thyroid hormone is replaced.

Along with its regulation by TSH, the thyroid gland also receives direct neuronal inputs from the autonomic nervous system. Fibers from the sympathetic and parasympathetic nervous systems penetrate the thyroid gland and terminate around arterioles, capillaries, and, less frequently, venules near follicular cells, providing additional mechanisms for regulation of thyroid hormone output.

RELEASE

The pituitary hormone TSH, and presumably the hypothalamic hormone TRH, is released in a pulsatile manner, with pulses occurring at approximately 90- to 180-min intervals in humans (Gore, 1998). Studies on the pulsatile release of TSH in humans demonstrate that its levels also change across the 24-h period, peaking from the evening through 0400 h, decreasing through

noon, and remaining low until the beginning of the next cycle in the evening. The increase in TSH release during the evening is referred to as the nocturnal TSH surge. This appears to be a true circadian rhythm, and it is driven by the suprachiasmatic nucleus of the hypothalamus (see Chapter 39).

It is important to note the terminology of *free* versus *bound* hormone. When the thyroid hormones are released into the circulation from the thyroid gland, a large portion bind to binding globulins. The free (or not protein-bound) thyroid hormone constitutes only <1% of all thyroid hormone circulating in the bloodstream. It is the free thyroid hormone that is responsible for the negative feedback regulation of TRH neurons and thyrotropes, as well as for actions at target tissues. Bound hormone is not immediately bioavailable but may provide a reservoir for thyroid hormones when they are needed.

Thyroid Hormones and Their Receptors

The two major thyroid hormones are T_4 and T_3, which have four or three iodide molecules, respectively (Fig. 38.4). T_3 has greater biological activity. Thyroid hormones contain two phenolic rings, which make these molecules lipophilic, facilitating their transport through the lipophilic membranes of target cells. Some plasma membrane transport molecules assist with this process because the thyroid hormones possess a polar alanine side-chain that appears to impede the process of passive diffusion. Thyroid hormones bind to the thyroid hormone receptors in target tissues. They also mediate negative feedback effects on TRH cells in the hypothalamus and pituitary. Although T_4 is released from the thyroid gland in greater quantities than T_3, most T_4 is converted to T_3, as catalyzed by two enzymes, type I and type II deiodinase. Type I deiodinase is found mostly in peripheral tissues and the pituitary gland, and type II deiodinase is found in the brain, pituitary gland, and brown fat. Because an important component in the synthesis of thyroid hormones is iodine, an iodine-deficient diet results in the impairment of thyroid hormone metabolism, leading to developmental abnormalities or adult dysfunctions.

With respect to the negative feedback regulation by thyroid hormone (Fig. 38.4), thyroid hormone receptors have been localized within TRH neurons in the paraventricular nucleus (Segerson et al., 1987). Thus, these actions can be mediated directly on TRH neurons in the brain. In addition, the expression of the thyroid hormone receptor is abundant in the anterior pituitary gland, particularly thyrotrope cells, indicating direct feedback action of the thyroid hormone on this organ as well.

The thyroid hormone receptor is a member of the nuclear hormone superfamily, with isoforms (α and β) and splice variants, most commonly α1 and β2 in the hypothalamus (Lechan, Qi, Jackson, & Mahdavi, 1994). These are nuclear transcription factors that bind to a regulatory element on the DNA to activate or inhibit gene transcription. Binding of the thyroid hormone to its receptor causes the heterodimerization of the receptors with another member of the nuclear receptor superfamily, RXR, and facilitates the binding of the dimerized receptor-ligand complex to the promoter of the target gene, thereby modulating gene transcription.

Disruptions in the Thyrotropic Axis

Thyroid hormone excess or insufficiency can have profound effects on the central nervous system, particularly during development. Thus, a normal ("euthyroid") tone is necessary for the health and survival of the organism. Excessive thyroid hormone, or hyperthyroidism, can result in symptoms such as tremor, nervousness, tremulousness, insomnia and sleep disorders, and impairments in memory and concentration. More profound effects include neuropsychiatric syndromes, seizure, chorea, and even coma. Deficiencies of the thyroid hormone, called hypothyroidism, can cause problems such as sleep apnea, hypothermia, hypoventilation, neuropsychiatric syndromes, peripheral neuropathy, cerebellar ataxia, and coma. During chronic thyroid hormone deficiency, fluids can accumulate and result in puffiness of the hands, feet, face, and tongue and cause speech impairment. This is called myxedema. Hypothyroidism generally is treated by thyroid hormone replacement.

When an organism is exposed to insufficient thyroid hormone during fetal or early neonatal development, it can develop cretinism, characterized by severe deficits in mental, neurological, and physical development. Most of these deficiencies can be mitigated or even reversed by newborn infant screening and thyroid hormone replacement.

Hypothalamic Control of Growth

Overview, Characteristics, and Physiological Functions of the Hypothalamic-Somatotropic Axis

The somatotropic axis controlling body growth has three levels: the primary hypothalamic hormones, of which there are two, growth hormone-releasing hormone (GHRH) and a growth hormone-inhibiting hormone, called somatostatin; the secondary pituitary hormone, called growth hormone (GH) or somatotropin; and the tertiary target organ—in this case the liver—which produces insulin-like growth factor I (IGF-I) in response to the GH stimulus (Fig. 38.5). A special characteristic of the somatotropic axis is that both its pituitary hormone, GH, as well as GH's target hormone, IGF-I, can exert growth-promoting effects alone or in combination. In fact, GH and IGF-I act synergistically with each other to affect somatic growth. However, both of these hormones play important roles, as mutations

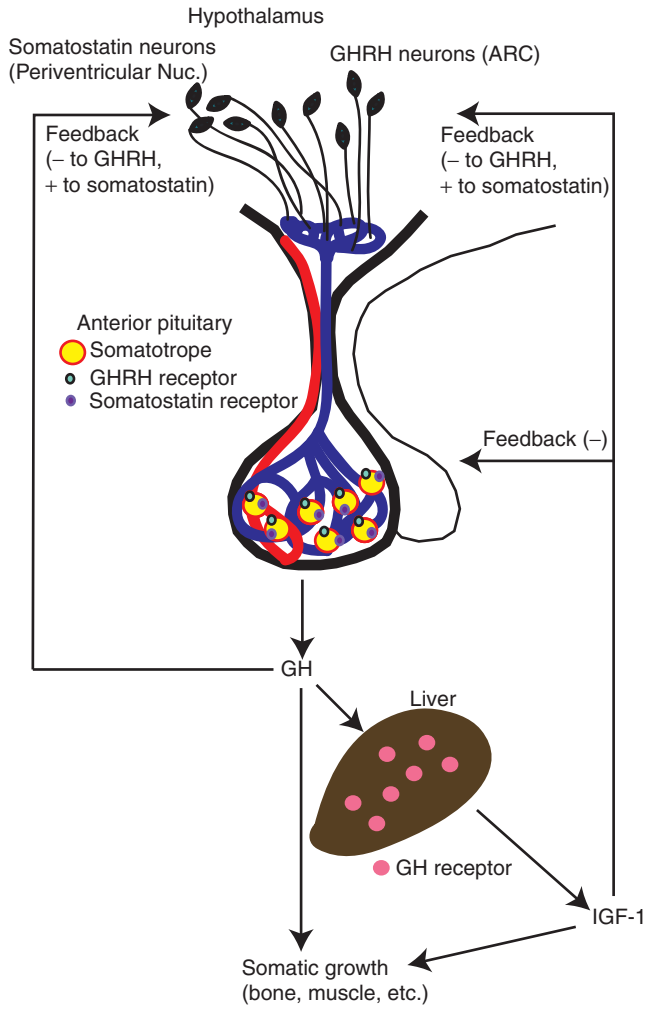


FIGURE 38.5 The hypothalamic control of somatic growth is shown. GHRH neurons in the arcuate nucleus, and somatostatin neurons in the periventricular region of the anterior hypothalamus, provide the stimulatory and inhibitory inputs to pituitary somatotropes, respectively (via the portal capillary vasculature). The synthesis and release of GH is stimulated (GHRH) or inhibited (somatostatin) at the somatotropes. This molecule acts at the liver to cause the biosynthesis and release of IGF-I. Both GH and IGF-I are released into the general circulation to affect somatic growth, and they both feed back negatively at the somatotropes of the pituitary and the GHRH neurons of the arcuate nucleus. IGF-I and growth hormone also feed back to stimulate somatostatin. Abbreviations: ARC, arcuate nucleus; GH, growth hormone; GHRH, growth hormone-releasing hormone; IGF-I, insulin-like growth factor I.

resulting in the absence of GH, IGF-I, or their receptors result in dwarfism.

Unlike most other neuroendocrine systems, the somatotropic axis has two known hypothalamic hormones that play important but opposing roles (Fig. 38.5). One hormone, GHRH, is stimulatory to the release of the pituitary GH. GHRH somata are localized in the arcuate nucleus of the hypothalamus. The other hypothalamic hormone, somatostatin, inhibits GH release. Perikarya of the somatostatin neurons involved in the regulation of growth hormone release are found primarily in the periventricular region of the anterior hypothalamus. Although arising from different sources, these hormones are both released into the portal capillary vasculature to affect the synthesis and release of GH, from specialized cells in the anterior pituitary gland called somatotropes, constituting approximately 30% of pituitary cells. The GH binds to its receptors in the liver to stimulate the release of the tertiary hormone, IGF-I. Both GH and IGF-I have somatotropic actions on target organs and tissues undergoing growth, regeneration, or metabolic homeostasis, such as bone and muscle.

PHYSIOLOGY

GH plays its major role in neonatal and postnatal growth and is particularly important in the pubertal growth spurt. Its anabolic activities are critical for metabolic homeostasis. The growth hormone stimulates protein synthesis, influences carbohydrate metabolism, enhances glucose uptake and utilization, and has lipolytic effects. At the systems level, it stimulates the growth of organs, such as the heart and kidney, and plays a role in hormone production, skeletal growth and maturation, and immune function. Many of its effects are sexually dimorphic, as are the distribution of its receptors and other proteins involved in this axis.

Growth hormone and IGF-I are both involved in the feedback regulation of the somatotropic axis (Fig. 38.5). In the case of GHRH, GH acts as a classical negative feedback molecule, downregulating the synthesis and secretion of GHRH. By contrast, GH *stimulates* gene expression of the inhibitory hypothalamic hormone somatostatin. In the absence of GH, somatostatin mRNA levels are downregulated. IGF-I also acts as a feedback regulator of the somatotropic axis, negatively regulating GHRH levels and stimulating somatostatin gene expression.

RELEASE

Patterns of somatotropic hormones are similar to those for other neuroendocrine systems. Both GHRH and somatostatin are released from hypothalamic nerve terminals into the portal circulation in a pulsatile manner, similar to other neurohormones. Importantly, the growth hormone is released in a strongly diurnal pattern and its release is associated with sleep, with highest levels occurring shortly after the onset of slow-wave sleep, and most GH secretion occurring at night in humans.

Growth Hormone, IGF-I, Binding Proteins, and Receptors

GROWTH HORMONE

Growth hormone, or somatotropin, is a 191–amino acid protein that contains two disulfide bonds that give it a

three-dimensional conformation that is necessary for its biological activity. Release of the growth hormone is triggered by GHRH and is inhibited by somatostatin. The growth hormone plays a negative feedback role in the somatotropic axis, feeding back at the level of the GHRH neuron in the hypothalamus to inhibit GHRH mRNA levels and synthesis and stimulating somatostatin gene expression and release. Not only does the regulation of growth hormone involve the balance between GHRH and somatostatin, but it also involves other neurotransmitters and neuroactive substances, such as dopamine, vasoactive intestinal peptide, pituitary adenylate cyclase-activating peptide, TRH, galanin, neuropeptide Y, motilin, and interleukins. In addition, other hormones, such as the thyroid hormone, glucocorticoids, sex steroid hormones, and metabolic fuels, can influence the release of growth hormone, highlighting the interwoven nature of the different neuroendocrine systems.

The growth hormone is released into the general circulation, where approximately 50% of this molecule is transported in association with a growth hormone–binding protein that is structurally similar to the extracellular domain of the growth hormone receptor. This molecule and the growth hormone receptor (GH-R) are alternative splice products of the same precursor RNA: one encoding the growth hormone–binding protein and the other the GH-R.

IGF-I

A major target for the growth hormone is the IGF-I system. IGF-I is a growth factor that exerts effects on neurite outgrowth and nervous system development. In addition, it is an important mediator of the somatotropic axis and, in fact, was originally referred to as somatomedin. Both GH and IGF-I individually have similar effects on somatic growth and development; however, the two hormones act much more efficiently in combination. In addition, it appears that the growth-promoting effects of IGF-I are particularly important in the central nervous system. This is consistent with the observation that along with its production by the liver, IGF-I is also produced in the brain, and it is probably this central IGF-I that is responsible for many of its neurotrophic effects.

Like the growth hormone, IGF-I plays a feedback role in the inhibition of GHRH biosynthesis (Fig. 38.5). This may involve central and not peripheral IGF-I, as most animal studies have demonstrated that only centrally but not peripherally administered IGF-I has this effect. At the level of the pituitary, IGF-I appears to be the major regulator of negative feedback of growth hormone synthesis and release. Somatotropes contain IGF-I receptors, and IGF-I inhibits growth hormone gene expression and release under basal and GHRH-stimulated conditions. IGF-I can also act as a negative feedback regulator of GHRH neurons, as it can cross the blood–brain barrier to a limited extent. Moreover, IGF-I is synthesized in the central nervous system, including the hypothalamus, and therefore local IGF-I can exert an influence on GHRH neurons. It also has been shown that IGF-I can stimulate somatostatin release, albeit at higher doses than those necessary to inhibit GHRH release, suggesting that IGF-I may be more involved in the suppression of GHRH than the enhancement of somatostatin levels.

IGF-I in the circulation binds to several binding proteins, the major one being IGF-I binding protein 3 (IGFBP-3). Free IGF-I is more biologically active than bound IGF-I, and therefore IGFBP-3 levels are an important regulator of the actions of IGF-I.

Disruptions of the Somatotropic Axis

Somatic growth and development are affected by disruptions of the somatotropic axis, and the consequences are dependent on when during development the system is affected. In early development, growth hormone deficiency can result in an extremely short stature, with adults sometimes reaching a maximal height of less than 1 m. Growth hormone deficiency is generally treated with growth hormone replacement using recombinant growth hormone. In some cases, GHRH, rather than growth hormone, is replaced in a pulsatile manner. Dwarfism is sometimes caused by mutations in the IGF-I system, or in the GH receptor. Laron dwarfism, characterized by mutations in the GH receptor gene, results in very low IGF-I levels, but GH levels are high, probably due to reduced negative feedback from the low IGF-I concentrations.

If GH is oversecreted or dysregulated during development, the consequence can be gigantism, resulting in an extremely large stature. When abnormal GH stimulation occurs after adulthood, this results in acromegaly. In adulthood, only certain tissues remain responsive to growth hormone stimulation, including the jaws, nose, orbital ridges, fingers, elbows, and knees, and these tissues are stimulated to grow, while other tissues are not, causing deformations in proportions and appearance.

Hypothalamic Control of Reproduction

Overview, Characteristics, and Physiological Function of the Hypothalamic-Pituitary-Gonadal Axis

Reproduction is critical to the survival of the species. In vertebrate organisms, the hypothalamic-pituitary-gonadal (HPG) axis hierarchical organization includes (1) the gonadotropin-releasing hormone (GnRH) neurons of the hypothalamus and preoptic area; (2) the gonadotropes of the anterior pituitary gland; and (3) the gonads (ovary in females and testis in males) (Fig. 38.6). The GnRH peptide, cell bodies of which are localized in the preoptic area and mediobasal hypothalamus, is released

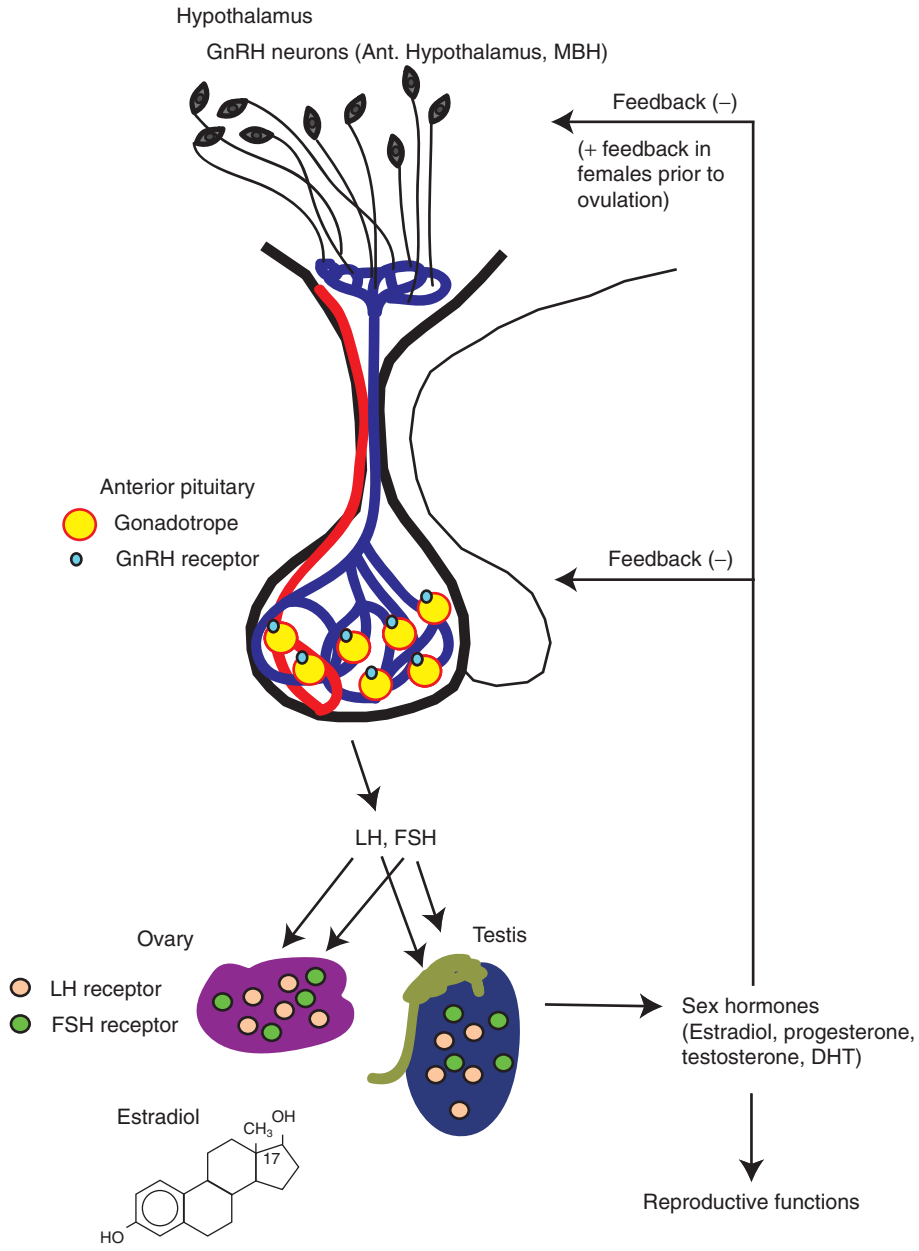


FIGURE 38.6 The hypothalamic–pituitary–gonadal axis controlling reproduction is shown. GnRH neurons in the preoptic area/anterior hypothalamus (rodents) or basal hypothalamus (primates) project their neuroterminals to the external zone of the median eminence. GnRH travels through the portal vasculature to cause the synthesis and release of the gonadotropins, LH and FSH, from gonadotropes of the anterior pituitary. These molecules act at the gonads (ovary and testis) to cause the biosynthesis and release of sex steroid hormones. These are released into general circulation to affect secondary sex characteristics and act at other somatic tissues. Sex steroid hormones feed back negatively at the brain and pituitary to regulate biosynthesis of GnRH and the gonadotropins in males and females. In addition, prior to ovulation, these steroids exert positive feedback in females. The structure of estradiol is shown (bottom). Abbreviations: FSH, follicle-stimulating hormone; GnRH, gonadotropin-releasing hormone; LH, luteinizing hormone; MBH, medial basal hypothalamus.

from neuroterminals in the median eminence in a pulsatile manner, acting upon the pituitary gonadotropes to cause pulsatile release of the pituitary gonadotropins: luteinizing hormone (LH) and follicle-stimulating hormone (FSH). These hormones in turn bind to receptors on specialized cells in the ovaries or testes to stimulate steroidogenesis, spermatogenesis in males, and oogenesis in females. The gonads of both sexes respond to this activation by synthesizing and secreting the sex steroid hormones, estrogens, progestins, and androgens. Although the dominant steroid hormones produced by the ovaries are estrogens and progestins, ovaries also produce appreciable amounts of androgens. The testes produce androgens at higher levels than females, but they also secrete estrogens and progestins. Thus, there is no such thing as a "male" or "female" sex steroid hormone. Rather, what differs between the sexes are the concentrations of these hormones.

Unlike other neuroendocrine cells, GnRH neurons of most vertebrates have an unusual embryonic origin, deriving from cells that are first detected in the olfactory placode (Wray, 2001). During embryonic development, GnRH neurons migrate into the brain, reaching their final destinations in the preoptic area and hypothalamic regions during mid- to late embryogenesis. Also, unlike most other neuroendocrine peptide systems, GnRH neurons are relatively scattered through these regions and are not tightly concentrated in a single discrete nucleus.

GnRH neurons are functionally regulated by numerous neurotransmitters and neurotrophic factors, and this is probably related to the integrative nature of the GnRH neuron (Gore, 2002). In order for reproduction to be successful, the organism must be able to assess the conditions of internal and external environment to determine that conditions are optimal. The numerous inputs to GnRH neurons from other sensory and neuroendocrine cells enables these key reproductive neurons to evaluate the organism's health and fitness and the external conditions (e.g., season, food availability, the presence of predators or a conspecific mate) and to coordinate the timing of reproduction, which is quite costly, with the optimal external and internal circumstances. Thus, GnRH cells are regulated by inputs from neurotransmitters such as glutamate, GABA, catecholamines and monoamines, neuropeptides such as kisspeptin, CRH, β-endorphin, vasopressin, neurotensin, neurotrophic factors, and many other substances. In this way, GnRH neurons receive cues regarding nutritional status, thermogenesis and exercise, stress, and social cues and can integrate these inputs to coordinate the timing of reproduction with the appropriate stimuli.

PHYSIOLOGY

The sex steroid hormones, once released into the circulatory system, act at receptors that are located throughout the body to exert their classical effects on secondary sex characteristic development and on adult reproductive function. In addition, sex steroid hormones act on receptors in the brain, including the hypothalamus, to exert diverse effects on the CNS. One of the major effects of sex steroid hormones is as a feedback modulator of the hypothalamic GnRH neurons. However, along with the traditional negative feedback exerted in neuroendocrine systems, in females, ovulation is controlled by *positive* feedback from sex steroid hormones. This concept of negative/positive feedback regulation is discussed later, and is unique to the reproductive neuroendocrine system of females.

Not surprisingly, functions of the reproductive axis change during the life cycle of the organism, as reproductive function varies with developmental stage (Plant, 1988). During the late embryonic/early postnatal period, as the GnRH system finishes its neuroanatomical development, there is increased activity of the brain–pituitary–testicular axis in males and, to a lesser extent, the brain–pituitary–ovarian axis in females. Shortly after birth, the reproductive system becomes quiescent, possibly due to the development of inhibitory inputs to GnRH neurons or the lack of stimulatory inputs to this system. This quiescent period, sometimes called the "prepubertal hiatus," is characterized by low GnRH, gonadotropin, and steroid hormone release.

At the onset of puberty, the GnRH system becomes activated, probably due to the removal of inhibitory afferents to GnRH neurons and the increasing dominance of excitatory inputs. Although the neural circuits controlling GnRH are complex, pharmacological studies suggest that the major inhibitory neurotransmitters putting the brake on GnRH prior to puberty are GABA and, to a lesser extent, endogenous opioid peptides and melatonin. As this inhibitory influence wanes, inputs from excitatory neurotransmitters begin to dominate, resulting in increased GnRH release. To date, numerous substances have been implicated in the activation of GnRH neurons at the onset of puberty, including glutamate (through both NMDA and non-NMDA glutamatergic receptors), kisspeptin, norepinephrine, and growth factors such as transforming growth factor-α and IGF-I. Once mature reproductive function has been attained (ovulation in females, viable sperm in ejaculates of males), the pulsatile release of GnRH and its regulation of the reproductive axis continues for much of the life cycle. In many species, females exhibit reproductive cycles (estrous cycles, menstrual cycles) of varying lengths, such as the 4-day estrous cycles of small rodents and the 28-day menstrual cycles of humans and many nonhuman primates.

Many species are sexually active during certain periods of the year, and as infrequently as once per year or even less. This seasonality enables organisms to coordinate the energetically demanding process of pregnancy, lactation, and rearing of the offspring with the appropriate environmental conditions. Seasonal breeding animals such as sheep have an almost completely inactive reproductive axis during the nonbreeding season, with very low GnRH release, low basal and GnRH-stimulated gonadotropin release, and a lack of ovulation. In many of these species, day length is what drives the seasonal breeding cycles, although this varies among the species. For hamsters, short nights stimulate the reproductive system, whereas short nights inhibit the reproductive axes of sheep. Males of some species undergo a process of testicular regression and recrudescence in coordination with the season. The pineal hormone melatonin appears to be critical to the coordination of reproduction with the season (Revel, Masson-Pevet, Pevet, Mikkelsen, & Simonneaux, 2009). It has been proposed that a "photoneuroendocrine" system exists whereby melatonin conveys signals about day length to the HPG axis, a process that involves input from the SCN (see Chapter 39) and convergence onto neurons that regulate GnRH release (Revel et al., 2009).

In human females but not most other species, menopause, the cessation of menstrual cycles and lack of ovulation, signals the end of the reproductive life cycle. The most obvious characteristic of menopause is a loss of ovarian follicles, called follicular atresia. Although many other mammals undergo reproductive senescence,

characterized by a lack of spontaneous reproductive cycles, this is not always due to follicular atresia. Instead, in these species, such as rodents, reproductive senescence is probably due to the loss of drive from hypothalamic GnRH neurons and/or changes in pituitary gonadotropin responses to GnRH rather than overt ovarian changes. In humans, it is quite possible that there are also hypothalamic and pituitary changes that may play a role in the menopausal process that are only just beginning to be understood.

RELEASE

A pulsatile manner of GnRH release is necessary for the maintenance of pituitary gonadotropin function. The pulsatile release of GnRH is more or less constant in males, with pulses occurring approximately every 1–2 h. However, in females, the nature of GnRH pulses varies across the reproductive cycle. In primates that have a 28-day menstrual cycle, LH pulses (measured as proxy for GnRH pulses) occur approximately once per hour during the early follicular phase, but slow to a frequency of once every 90 min later in the follicular phase as estrogen negative feedback increases. Just prior to ovulation, a large increase in GnRH release can be measured, although it is still controversial whether discrete pulses continue to be superimposed on this "GnRH surge." Nevertheless, it is likely that there is a change in the mode of GnRH release during the preovulatory period.

The pulsatile release pattern of GnRH, LH, and FSH is superimposed upon a diurnal pattern of the release of these hormones (Gore, 1998). Thus, the timing of the preovulatory LH surge in rodents is coordinated with the 24-h light/dark cycle (i.e., it is a diurnal but not a circadian rhythm). Although in women these diurnal rhythms are not as obvious as those in rats, it appears that the release of gonadotropins has a weak diurnal rhythm in humans. Men have also been reported to have nocturnal increases in reproductive hormone release.

Gonadal Steroid Hormones and Their Receptors

Sex steroid hormones are synthesized from the precursor cholesterol in a multistep process. Cholesterol first is converted to pregnenolone, which is converted to progesterone. This hormone is a major end point for biosynthesis in gonadal tissues and exerts potent effects on the body and brain. Progesterone serves as a precursor for the synthesis of testosterone following conversion first to 17α-hydroxyprogesterone and then androstenedione. Testosterone is the dominant steroid hormone in males, although it is also found in females. It serves as a precursor via the enzyme 5α-reductase for 5α-dihydrotestosterone, which is responsible for many of the masculinizing effects of steroids in males. Testosterone is also the direct precursor via the enzyme p450 aromatase for 17β-estradiol, which is the primary estrogen responsible for sexual behavior and physiology in females, and which also plays roles in masculine physiology and behavior. The sex steroid hormones bind to their receptors, members of the nuclear hormone superfamily. The binding of a steroid hormone to its receptor, and subsequent receptor dimerization, allows these transcription factors to bind to specific sites on the promoter of the target gene, called androgen-response elements (AREs), estrogen-response elements (EREs), and progesterone-response elements (PREs). This enables the receptor-ligand complexes to induce or repress gene transcription in their target tissues.

Sex steroid hormones are synthesized and secreted from the ovaries or testes to regulate the expression of secondary sex characteristics and to maintain adult reproductive function. In the Leydig cells of the testis, LH released from the pituitary gland is responsible for testosterone biosynthesis. Together with FSH, testosterone acts to coordinate and stimulate spermatogenesis in testicular seminiferous tubules. The testosterone that is released from the testis acts at secondary sex targets such as the internal and external genitalia (prostate, seminal vesicles, penis) to cause their development and maintain their functions. In addition, testosterone acts at muscle to increase muscle mass, causes the growth of facial and body hair, and enlarges the epiglottis, which is responsible for the deepening of the voice. In the ovary, FSH stimulates follicular development during each ovarian cycle, and LH is the primary stimulator of ovulation. Steroidogenesis in the ovary is also stimulated by FSH and LH. The resulting ovarian hormones, primarily 17β-estradiol and progesterone, are responsible for the development of female secondary sex characteristics such as the genitalia and mammary tissue.

Sex steroid hormone feedback in the male is accomplished by the usual neuroendocrine negative feedback mechanism (Fig. 38.6). Pulses of GnRH cause pulses of the gonadotropins, which in turn are responsible for pulses of testosterone. Feedback of testosterone onto the brain causes an inhibition of hypothalamic GnRH neurons, and testosterone also feeds back negatively onto the pituitary gonadotropes. At the pituitary gland, sex steroid hormones appear to act primarily to decrease the sensitivity of pituitary gonadotropes to GnRH, resulting in a decrease in LH pulse amplitude.

In the female, the ovarian steroid hormones estradiol and progesterone can act at the hypothalamus to provide negative feedback. During most of the reproductive cycle, these hormones feed back at the brain and the pituitary gland to suppress the release of GnRH and the gonadotropins, respectively. However, during the

mid- to late follicular stage in primates, or on the day of proestrus in rats with a four-to five-day estrous cycle, feedback effects of estrogens on GnRH release become positive. The stimulatory effects of estradiol on the hypothalamus result in the occurrence of the "preovulatory GnRH surge," which in turn triggers a preovulatory LH surge that triggers the final maturation of the ovum in the follicle and results in ovulation and the subsequent development of the corpus luteum. After ovulation, feedback of ovarian steroid hormones upon the hypothalamus and pituitary reverts to the negative mode until the next cycle.

Unlike the other neuroendocrine systems, the primary feedback site of gonadal steroid hormones does not appear to be the GnRH neurons. Although GnRH neurons express the ERβ, they do not express appreciable levels of ERα, androgen receptors (AR), or progesterone receptors (PR). Therefore, it is thought that other neural cells that make afferent inputs to GnRH neurons, and which express nuclear sex steroid receptors, may mediate effects of sex steroid feedback onto GnRH cells. The mediation of gonadal steroid feedback by intermediary cells in the CNS, primarily those expressing the ERα, may also help to explain the "switch" between negative and positive feedback onto the GnRH neurons in females (Wintermantel et al., 2006). For example, some afferent inputs to GnRH neurons that inhibit GnRH release may exert a dominant tone during most of the reproductive cycle and be responsible for negative feedback. Other afferent inputs that stimulate GnRH release may be regulated by higher levels of estrogens and underlie the positive feedback response to ovarian steroids that results in the preovulatory GnRH surge.

Disruptions in the Reproductive Axis

Mutations to the GnRH or GnRH receptor genes result in failure to undergo reproductive development and to attain normal adult maturation. More recently, mutations in other hypothalamic neurons that converge upon GnRH neurons have been associated with partial or incomplete reproductive capacity. Of these, kisspeptin and its receptor (GPR54) are particularly important, and emerging evidence shows that the ligands and receptors involved in fibroblast growth factor, neurokinin B, prokineticin, and other parts of the neural circuitry controlling GnRH release are involved (Caronia et al., 2011).

Other defects of the reproductive system involve the timing of the onset of puberty. In idiopathic central precocious puberty, GnRH is released far earlier than is normal. This results in activation of the release of pituitary gonadotropins, stimulation of the gonads, and the early development of secondary sexual characteristics. Treatment with long-lasting GnRH agonists, which initially stimulate but rapidly downregulate pituitary GnRH receptors, cause a downregulation of the reproductive axis and can halt pubertal development until the normal timing of puberty occurs. In the case of delayed puberty, the normal pubertal release of GnRH does not occur, and exogenous GnRH can be infused in a pulsatile manner to cause the stimulation of puberty in these individuals.

Hypothalamic Control of Lactation

Overview, Characteristics, and Physiology of the Hypothalamic-Lactational Axis

The pituitary hormone prolactin is a major regulator of lactation in mammals (Grattan & Kokay, 2008). However, unlike the other pituitary hormones, a unique releasing neuropeptide hormone for prolactin has not been identified in the hypothalamus, although there is a prolactin-inhibiting factor, dopamine (Fig. 38.7). Prolactin is synthesized in pituitary cells called lactotropes. Its release

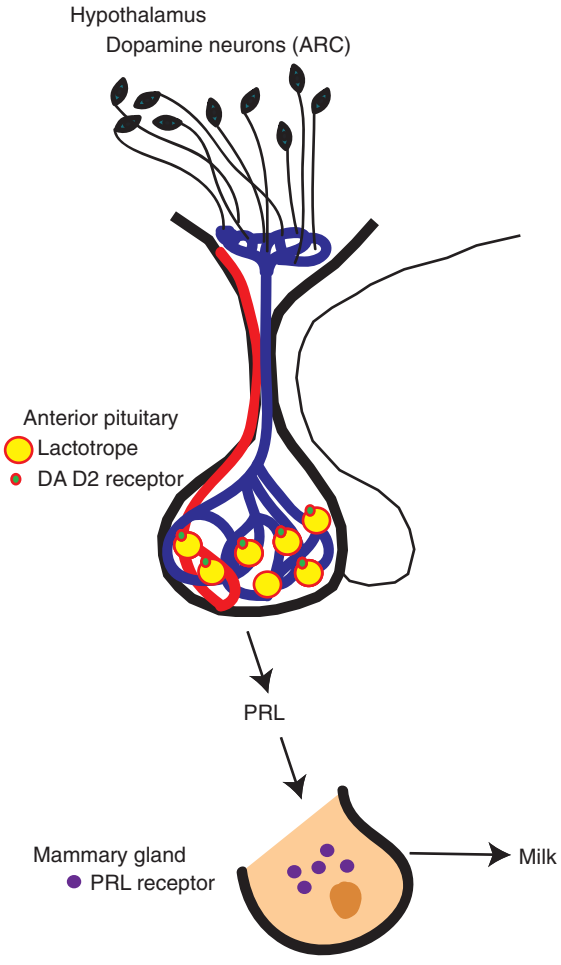


FIGURE 38.7 The hypothalamic control of lactation is depicted. Hypothalamic cells in the arcuate nucleus release dopamine into the portal capillary vasculature. The dopamine exerts negative regulation of prolactin release under most circumstances, via inhibitory D2 receptors. During lactation this inhibition is relieved and prolactin is released from the lactotropes of the anterior pituitary gland. Abbreviations: ARC, arcuate nucleus; DA, dopamine; PRL, prolactin.

into the circulatory system from the anterior pituitary gland causes milk synthesis and secretion into the alveoli of the mammary glands in response to the suckling stimulus. Prolactin, together with oxytocin, which is involved in milk ejection (see later), is responsible for the maintenance of lactation.

It is generally accepted that hypothalamic dopamine plays a major role in the inhibition of prolactin release, and dopamine is the major prolactin-inhibiting factor (Ben-Jonathan, Arbogast, & Hyde, 1989). Dopamine is abundant in the hypothalamus and is found in high concentrations in the median eminence, where it acts as an endocrine hormone to influence the function of the anterior pituitary gland via the portal capillary system. Dopamine receptors are present on pituitary lactotropes, and changes in dopamine concentrations in the portal capillary system correlate with changes in prolactin release under physiological conditions. The major hypothalamic population of dopamine neurons believed to regulate prolactin release is called the tuberoinfundibular dopaminergic system, also referred to as the A12 cell group. Their perikarya lie in the arcuate nucleus.

Regarding the stimulation of prolactin release by a hypothalamic factor(s), though many molecules have been proposed, none appears to be a unique prolactin-releasing hormone. TRH, vasoactive intestinal polypeptide, and oxytocin can stimulate prolactin release but they do not appear to have the one-to-one relationship between a true hypothalamic releasing factor and its adenohypophysial hormone. Thus, the search for a unique hypothalamic-releasing hormone for prolactin is ongoing.

PHYSIOLOGY

The release of prolactin varies by age, sex, and reproductive status. Levels of prolactin are higher in females than males, probably due to the higher estrogen levels in females, which stimulate prolactin release. During reproductive cycles in females, prolactin levels vary, with levels being higher during the luteal (postovulatory) than the follicular (preovulatory) phase in humans. In rats, prolactin levels peak shortly before the preovulatory LH surge that causes ovulation, a time when estrogen levels are high. Prolactin levels also increase during pregnancy, decrease just before birth, and increase again in response to suckling.

The role of prolactin in lactation is its best-studied and probably most important function. In all mammals, prolactin acts on the mammary gland to prepare and maintain its secretory activity. It causes the growth of ductal and lobuloalveolar tissues and causes increases in mRNA levels and translation of the milk protein, casein. Prolactin release is stimulated directly by suckling and results in the synthesis of milk. In addition, there is a relationship between suckling and the inhibition of reproductive function, as suckling can delay implantation of fetuses conceived during a postpartum estrus or cause long periods of anovulation. Prolactin may also play a role in maternal behavior, although this is not well understood.

Prolactin can influence the reproductive axis through its actions on the ovary. There, prolactin causes increases in LH receptors in the corpus luteum, progesterone synthesis by the corpus luteum, and high- and low-density lipoprotein binding in membranes of the corpus luteum. These functions may be related to uterine endometrial development and implantation of the blastocyst in rats, although it is not clear if this is the case in humans. Other functions of prolactin include osmoregulation, primarily in fish and amphibians.

RELEASE

Like the other pituitary hormones, prolactin is released from the anterior pituitary in a pulsatile manner. It has a diurnal rhythm, as it is driven by sleep cues rather than the circadian clock, similar to its related hormone, growth hormone (Gore, 1998).

Disruptions in the Prolactin System

Hyperprolactinemia is a pathological elevation in the secretion of prolactin. This is a common human pituitary disorder and can be a result of disrupted dopamine transmission, which normally would suppress prolactin release. Hyperprolactinemia can also result from pituitary adenomas that secrete prolactin. As a result of hyperprolactinemia, inappropriate lactation can occur, as can be the case with drugs that affect the dopamine system (neuroleptics, dopamine receptor blockers, antidepressants, antihypertensives, oral contraceptives). This syndrome can occur in both males and females.

HYPOTHALAMIC CONTROL OF POSTERIOR PITUITARY HORMONES AND THEIR REGULATED FUNCTIONS

Although the major focus of this chapter is the neuroendocrine regulation of anterior pituitary function, it is important to introduce the neuroendocrine control of neurohypophysial function. The posterior pituitary releases two major neurohormones, oxytocin and vasopressin (referred to here as arginine vasopressin or AVP; also called anti-diuretic hormone or ADH). The hypothalamic-neurohypophysial AVP and oxytocin neurons have large cell somata that are 20–40 μm in diameter, and are referred to as magnocellular neurons (as opposed to their parvocellular counterparts of 10–15 μm). These cells are localized in the supraoptic nucleus (SON), comprised entirely of magnocellular neurons; and the paraventricular nucleus (PVN), which has both magnocellular and parvocellular divisions (Burbach, van Schaick, de Bree, Lopes da Silva, & Adan, 1995). The hypothalamic-neurohypophysial somata

in the SON and PVN project axons that traverse the pituitary stalk and terminate in the posterior pituitary gland (Fig. 38.8). These axon terminals are abundant in secretory granules storing oxytocin or AVP, and in response to depolarization of their cell bodies, release these neuropeptides into general circulation.

PHYSIOLOGY

AVP and oxytocin have numerous functions in the brain and body, and a brief discussion of their diverse neuroendocrine-neurohypophysial functions is provided here (Burbach, Luckman, Murphy, & Gainer, 2001). One of AVP's critical roles is salt and water balance (Frank & Landgraf, 2008; see Chapter 37). In mammalian systems, changes in osmolality of extracellular fluids alter secretion of AVP, and to a lesser extent oxytocin, which can regulate salt and water intake through actions on kidney tubules and by exerting behavioral effects on drinking and salt appetite. AVP plays a further role as a vasoconstrictor in response to hemorrhage. As discussed earlier, AVP also is involved in the regulation of the stress response. Some of these effects may be mediated through transport of AVP to pituitary corticotropes via the portal capillary vasculature leading to the anterior pituitary gland. In addition, neurohypophysial actions of AVP on osmolality may indirectly modulate the brain–pituitary–adrenal axis.

Oxytocin is best known for its roles related to parturition and milk ejection in mammals (Russell, Leng, & Douglas, 2003). Oxytocin is important for uterine contraction during delivery. With regard to lactation, oxytocin responds to the suckling stimulus by stimulating contraction of the cells around the alveoli of the mammary gland to stimulate milk ejection. Oxytocin, together with AVP, plays key roles in social behaviors related to successful reproduction, such as maternal and paternal behaviors and pair bonding (Lim & Young, 2006).

RELEASE

An obvious difference between the neurophysial and adenohypophysial systems is that the former is a direct neural link between hypothalamus and pituitary, whereas the latter is mediated by the portal capillary vasculature. The neural innervation of posterior pituitary enables much more rapid regulation, triggered by changes in blood osmolarity (AVP) or the suckling stimulus (oxytocin), processes that require immediate responses.

GENES

The oxytocin and AVP are 9-amino acid neuropeptides synthesized from genes localized on the same chromosome, but they are in a "head to head" organization separated by a domain called the intergenic region. Therefore, these genes are transcribed in opposite directions (Fig. 38.8; Young & Gainer, 2003). The AVP

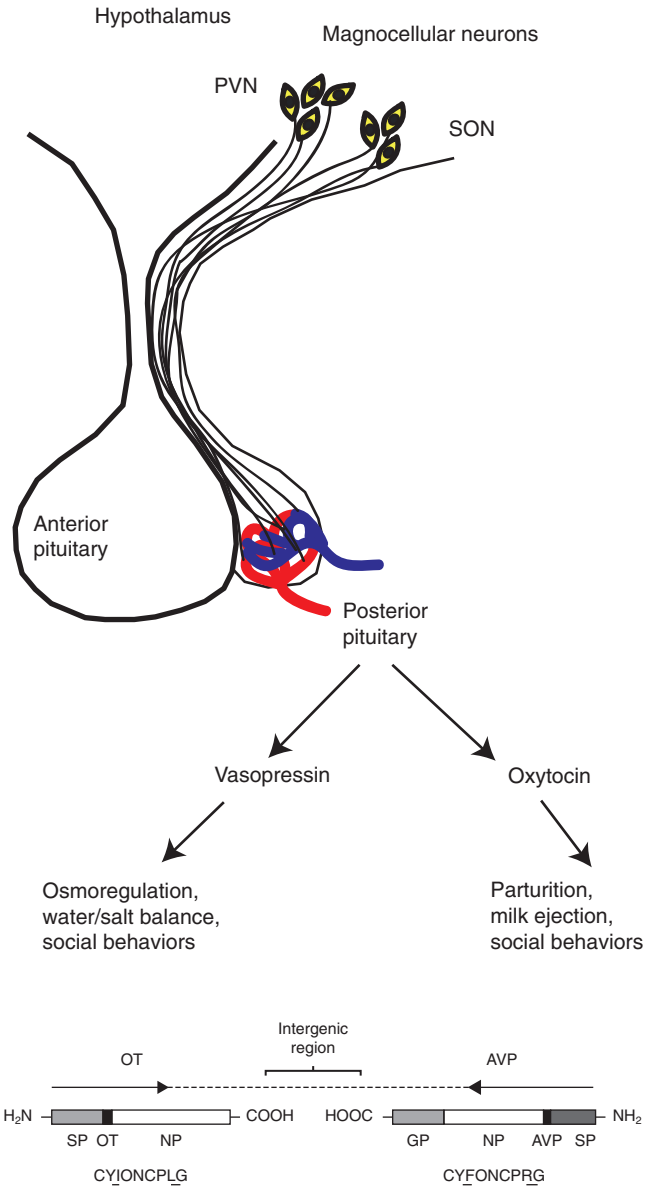


FIGURE 38.8 The hypothalamic-neurohypophysial control of vasopressin and oxytocin release is shown (top). Large cells (magnocellular) in the paraventricular and supraoptic nuclei synthesize vasopressin or oxytocin, which are transported by axonal transport into nerve terminals located in the posterior pituitary gland (neurohypophysis). When stimulated by depolarization, vasopressin and/or oxytocin are released from the posterior pituitary into general circulation. The structural organization of the vasopressin and oxytocin genes are shown (bottom). The RNA sequences are transcribed off of a single gene but in opposite directions, indicated by arrows. Each gene is made up of three exons that encode a signal peptide, the neurohormone (either vasopressin or oxytocin) and a neurophysin that may be involved in chaperoning the neuropeptides. Vasopressin also contains an additional sequence for a glycoprotein. The regions are separated by an intergenic sequence. The nonapeptide sequence of each neurohormone is shown below its respective genes, with differences indicated by underlines. Abbreviations: AVP, arginine vasopressin; GP, glycoprotein; NP, neurophysin; OT, oxytocin; PVN, paraventricular nucleus; SON, supraoptic nucleus; SP, signal peptide. For more information, see Young and Gainer (2003).

and oxytocin genes are structurally similar, comprising three exons that encode a signal peptide, the nonapeptide (either AVP or oxytocin), and a neurophysin whose function is not entirely known but may be involved in chaperoning the neuropeptides as they are processed in secretory vesicles (Fig. 38.8; Burbach et al., 2001).

HORMONES, THE BRAIN, AND BEHAVIOR

Central neuroendocrine control begins in the brain and ends with a target hormone or hormones exerting peripheral actions. At the same time, there is feedback to the CNS to enable adjustments in hypothalamic output. While the mechanisms for these regulatory loops differ from system to system, many of the neuroendocrine systems described above involve actions of steroid hormones on the nervous system. For example, HPG, HPA, and HPT hypothalamic hormones are regulated by gonadal, adrenal, and thyroid hormone feedback, respectively. This underscores the point that there are steroid hormone receptors in the brain mediating these actions of peripheral hormones. Furthermore, beyond the regulation of neuroendocrine physiology, peripheral hormones can also regulate and coordinate the behaviors connected to those physiological functions. The best example of this is the control of reproductive physiology and behavior by the gonadal steroids estrogens, progestins, and androgens. These hormones act upon estrogen receptors (ERs), androgen receptors (ARs), and progesterone receptors (PRs), beginning as early as fetal development (Roselli, Resko, & Stormshak, 2006). All of these receptors are abundant in the hypothalamus, and neuroscientists have begun to look beyond the hypothalamus as a CNS target for steroid hormones, in order to understand how these molecules may mediate or mitigate nonreproductive CNS functions including neuroprotection, synaptic plasticity, neurogenesis, cognition, and neurodegeneration.

Sexual Differentiation of the Brain

The brain is sexually differentiated, with morphological, anatomical, and functional differences between males and females (Schwarz & McCarthy, 2008). In most mammals and many other vertebrates, these differences appear to be established, or "organized," in late gestation and early postnatal life, due to early hormonal exposures. The developing fetus is exposed not only to its own gonadal hormones, but also to maternal and environmental factors, exposures to all of which have permanent effects on brain development, and ultimately, on adult neural and behavioral functions. Recent research suggests that the mechanisms by which hormones sculpt the developing nervous system involve epigenetic actions: nongenomic actions that affect gene regulation independent of changes in the DNA base pair sequence (McCarthy et al., 2009). Perinatal differences in sex steroid hormones act to alter the size of these brain regions such that they differ between males and females; in general, because the fetal testis is much more active than the ovary, testicular hormones are responsible for these neuroanatomical changes. Although it might be predicted that testosterone is responsible for these effects in males, depending upon the brain region, it is often testosterone's metabolite, 17β-estradiol, that is much more potent in many of these actions in mammalian species. Testosterone is a direct precursor to estradiol, metabolized by the P450 aromatase enzyme. This enzyme is found in the hypothalamus, as well as many other brain regions that are steroid sensitive. The perinatal *organizational* effects of testosterone and estradiol are crucial in order for normal sex-specific reproductive physiology and behavior to occur in adulthood. These behaviors also require further actions of gonadal steroid hormones, which are secreted in increasing amounts during the pubertal process. These so-called *activational* effects of steroids at puberty require that normal organization has occurred early in life. It is also becoming clear that the pubertal period is characterized by continued development of the brain, including neural plasticity that is modulated by steroid hormones. Thus, puberty may be considered as a second organizational period, as well as one during which activation of neonatally-organized circuits occurs (Sisk & Foster, 2004).

The issue of sexual dimorphism in the human brain has been quite controversial, but there is strong evidence for morphological and functional differences in specific brain nuclei in humans (Cahill, 2006). Sex differences in hypothalamic subnuclei, along with regions of the corpus callossum, hippocampus, amygdala, and neocortex have robust sex differences in size and neurochemistry, a difference speculated to be due to differential exposure to sex steroid hormones (Cahill, 2006).

Hormones and Sexual Behavior

There are clearly differences in the sexual behavior of males and females, effects that are dependent upon organizational and activational effects of gonadal steroid hormones. For example, copulatory behavior in male rodents comprises mounts of the female by the male, penile intromissions, and ejaculations. A receptive female rodent will respond to male copulatory behaviors with a lordosis response, characterized by a rigid stance and an arched back. These behaviors are driven in large part by the responses of hypothalamic nuclei to hormones, and the nature of the adult behavior is determined by the prenatal/neonatal exposure to hormones. Thus, the hypothalamus of adult males, which has been masculinized neonatally by testicular testosterone (or its aromatization product, estradiol), will respond to the testosterone produced by

the adult gonad (metabolized to DHT) by causing the typical male pattern of sexual behavior. In contrast, the female hypothalamus, which is not exposed neonatally to sufficient levels of ovarian hormones to cause masculinization of her brain, responds to estrogen produced by her adult ovary with the appropriate female receptive response.

The regions of the brain responsible for sex behavior differ between males and females. Sexual behavior in male rodents is largely driven by the preoptic area (POA) of the brain, just rostral to the hypothalamus, although other regions contribute to this process. Lesions of the POA in males cause the loss of sexual behavior, and testosterone implanted into the POA can induce sex behavior. In female rodents, the primary hypothalamic site regulating receptive sexual behavior is the ventromedial nucleus of the hypothalamus, although as in males, other regions also play contributing roles. Estradiol, followed by progesterone, facilitates the expression of the lordosis reflex and coordinates the timing of female receptivity with the preovulatory GnRH/LH surge that causes ovulation. Thus, the timing of mating behavior and ovulation is coordinated by the release of ovarian steroid hormones. Female reproductive cyclicity and the ability to ovulate in adulthood is also organized by sex-specific exposures to neonatal steroid hormones. This exemplifies the point that the timing of hormonal exposures has long-term consequences on both physiology and behavior in adulthood. Disruptions of endogenous hormones, or perturbations by exogenous hormonally-active substances (e.g., endocrine-disrupting chemicals) during critical developmental periods can have lifelong effects on reproductive behavior (Box 38.2).

Pheromones, the Vomeronasal Organ, and Sexual Behavior

Reproductive behavior in many species involves signals from pheromones, which are chemical signals used for communication between animals (Tirindelli, Dibattista, Pifferi, & Menini, 2009). Pheromones are chemical messengers that typically are produced by specialized glands, such as the sebaceous glands in the flanks of animals, or released in urine or vaginal fluid. They are released not into the body but rather into the external environment to convey information about the physiological status of the animal to other individuals.

The ability to detect pheromones in most mammals or other vertebrates involves the vomeronasal organ, an auxiliary olfactory organ found near the nasal cavity, and its projections to the accessory olfactory bulbs. Cells in these latter regions contact a number of brain regions, including the medial amygdala, cortical regions, and the bed nucleus of the stria terminalis, all of which are implicated in sex behavior. These regions in turn project to the hypothalamus and preoptic area. Thus, information from pheromones about mating status can be conveyed to parts of the brain coordinating reproductive physiology and behavior.

In most species, information about reproductive status, dominance status, or other social cues can be conveyed by pheromones, although it is still controversial whether humans have a vomeronasal organ and/or can sense pheromones. Those organisms using pheromones as chemosensory cues have developed mechanisms for depositing or releasing pheromones to "advertise" sexual competency or receptivity in order to attract a mate or to mark territory, to protect food supplies, mates, or offspring. Pheromones and the systems that mediate their effects play broader roles than those involved in reproductive physiology and behavior. Pheromones can dramatically influence the timing of sexual development and function. In female mice, exposure to male urinary pheromones accelerates the timing of pubertal development. Males respond to female urinary pheromones with sharp increases in luteinizing hormone release. It has also been reported that pheromones can coordinate the timing of reproductive cycles, even in humans, although the mechanisms and functions of these pheromones are still

BOX 38.2

ENDOCRINE-DISRUPTING CHEMICALS AND THEIR EFFECTS ON NEUROENDOCRINE SYSTEMS

Endocrine-disrupting chemicals (EDCs) are compounds in the environment that perturb endocrine systems. Recent interest has turned to whether, and how, EDCs might affect neuroendocrine function. Early life developmental periods—fetal, infant, and pubertal—are of particular interest because during these life stages, levels of endogenous hormones vary precipitously and are sex-specific. Thus, exposures of developing organisms to even very low levels of EDCs may have permanent organizational effects on the neuroendocrine system, affecting physiology and behavior in adulthood (Gore, 2008). Current research in the field is focusing on the molecular mechanisms by which developmental EDC exposure may "reprogram" the expression and regulation of neural steroid hormone receptors, including epigenetic mechanisms including altered DNA methylation and histone modifications.

Andrea C. Gore

controversial and poorly understood. Nevertheless, it is clear that pheromones play important roles in coordinating physiology and behavior.

References

Aguilera, G. (2011). HPA axis responsiveness to stress: Implications for healthy aging. *Experimental Gerontology, 46,* 90–95.

Ben-Jonathan, N., Arbogast, L. A., & Hyde, J. F. (1989). Neuroendocrine [corrected] regulation of prolactin release. *Progress in Neurobiology, 33,* 399–447.

Burbach, J. P. H., Luckman, S. M., Murphy, D., & Gainer, H. (2001). Gene regulation in the magnocellular hypothalamo-neurohypophysial system. *Physiological Reviews, 81,* 1197–1267.

Burbach, J. P., van Schaick, H. S., de Bree, F. M., Lopes da Silva, S., & Adan, R. A. (1995). Functional domains in the oxytocin gene for regulation of expression and biosynthesis of gene products. *Advances in Experimental Medicine and Biology, 395,* 9–21.

Cahill, L. (2006). Why sex matters for neuroscience. *Nature Reviews Neuroscience, 7,* 477–484.

Caronia, L. M., Martin, C., Welt, C. K., Sykiotis, G. P., Quinton, R., Thambundit, A., … Nelly Pitteloud. (2011). A genetic basis for functional hypothalamic amenorrhea. *NEJM, 364,* 215–225.

Frank, E., & Landgraf, R. (2008). The vasopressin system—from antidiuresis to psychopathology. *European Journal of Pharmacology, 583,* 226–242.

Gore, A. C. (1998). Circadian rhythms during aging. In P. R. Hof & C. V. Mobbs (Eds.), *The Endocrinology of Aging* (pp. 127–165). Basel: Karger Press.

Gore, A. C. (2002). *GnRH: The Master Molecule of Reproduction*. Norwell, MA: Kluwer Academic Publishers.

Gore, A. C. (2008). Developmental programming and endocrine disruptor effects on reproductive neuroendocrine systems. *Frontiers in Neuroendocrinology, 29,* 358–374.

Grattan, D. R., & Kokay, I. C. (2008). Prolactin: A pleiotropic neuroendocrine hormone. *Journal of Neuroendocrinology, 20,* 752–763.

Harris, G. W. (1971). Humours and hormones. *Journal of Endocrinology, 53,* ii–xxiii.

Kriegsfeld, L., Mei, D., Bentley, G., Ubuka, T., Mason, A., Inoue, K., … Silver, R. (2006). Identification and characterization of a gonadotropin-inhibitory system in the brains of mammals. *Proceedings of the National Academy of Sciences of the United States of America, 103,* 2410–2415.

Krsmanovic, L. Z., Hu, L., Leung, P. K., Feng, H., & Catt, K. J. (2009). The hypothalamic GnRH pulse generator: Multiple regulatory mechanisms. *Trends in Endocrinology and Metabolism, 20,* 402–408.

Lechan, R. M., Qi, Y., Jackson, I. M., & Mahdavi, V. (1994). Identification of thyroid hormone receptor isoforms in thyrotropin-releasing hormone neurons of the hypothalamic paraventricular nucleus. *Endocrinology, 135,* 92–100.

Lim, M. M., & Young, L. J. (2006). Neuropeptidergic regulation of affiliative behavior and social bonding in animals. *Hormones and Behavior, 50,* 506–517.

McCarthy, M. M., Auger, A. P., Bale, T. L., De Vries, G. J., Dunn, G. A., Forger, N. G., … Wilson, M. E. (2009). The epigenetics of sex differences in the brain. *Journal of Neurosciences, 29,* 12815–12823.

Morimoto, M., Morita, N., Ozawa, H., Yokoyama, K., & Kawata, M. (1996). Distribution of glucocorticoid receptor immunoreactivity and mRNA in the rat brain: An immunohistochemical and in situ hybridization study. *Neuroscience Research, 26,* 235–269.

Nillni, E. A. (2010). Regulation of the hypothalamic thyrotropin releasing hormone (TRH) neuron by neuronal and peripheral inputs. *Frontiers in Neuroendocrinology, 31,* 134–156.

O'Connor, T. M., O'Halloran, D. J., & Shanahan, F. (2000). The stress response and the hypothalamic-pituitary-adrenal axis: From molecule to malancholia. *QJM, 93,* 323–333.

Plant, T. M. (1988). Puberty in primates. In Knobil, E. & Neill, J. (Eds.), *The physiology of reproduction* (pp. 1763–1788). New York: Raven Press, Ltd.

Revel, F. G., Masson-Pevet, M., Pevet, P., Mikkelsen, J. D., & Simonneaux, V. (2009). Melatonin controls seasonal breeding by a network of hypothalamic targets. *Neuroendocrinology, 90,* 1–14.

Roselli, C. E., Resko, J. A., & Stormshak, F. (2006). Expression of steroid hormone receptors in the fetal sheep brain during the critical period for sexual differentiation. *Brain Research, 1110,* 76–80.

Russell, J. A., Leng, G., & Douglas, A. J. (2003). The magnocellular oxytocin system, the fount of maternity: Adaptations in pregnancy. *Frontiers in Neuroendocrinology, 24,* 27–61.

Schwarz, J. M., & McCarthy, M. M. (2008). Steroid-induced sexual differentiation of the developing brain: multiple pathways, one goal. *Journal of Neurochemistry, 105,* 1561–1572.

Segerson, T. P., Kauer, J., Wolfe, H., Mobitker, H., Wu, P., Jackson, I. M. D., & Lechan, R. M. (1987). Thyroid hormone regulates TRH biosynthesis in the paraventricular nucleus of the rat hypothalamus. *Science, 238,* 78–80.

Sisk, C. L., & Foster, D. L. (2004). The neural basis of puberty and adolescence. *Nature Neuroscience, 7,* 1040–1047.

Tirindelli, R., Dibattista, M., Pifferi, S., & Menini, A. (2009). From pheromones to behavior. *Physiological Reviews, 89,* 921–956.

Wintermantel, T. M., Campbell, R. E., Porteous, R., Bock, D., Grone, H. J., Todman, M. G., … Herbison, A. E. (2006). Definition of estrogen receptor pathway critical for estrogen positive feedback to gonadotropin-releasing hormone neurons and fertility. *Neuron, 52,* 271–280.

Wray, S. (2001). Development of luteinizing hormone releasing hormone neurones. *Journal of Neuroendocrinology, 13,* 3–11.

Yin, W., & Gore, A. C. (2010). The hypothalamic median eminence and its role in reproductive aging. *Annals of the New York Academy of Sciences, 1204,* 113–122.

Young, W. S. I., & Gainer, H. (2003). Transgenesis and the study of expression, cellular targeting and function of oxytocin, vasopressin and their receptors. *Neuroendocrinology, 78,* 185–203.

Andrea C. Gore

This page intentionally left blank

CHAPTER

39

Circadian Timekeeping

INTRODUCTION

Life on Earth has adapted to the cyclic variation in conditions imposed by our planet's daily rotation on its axis and its annual revolution around the sun. Almost all organisms studied, from cyanobacteria to *Homo sapiens*, have intrinsic time-keeping systems that adjust physiology and behavior to anticipate the daily cycles of light and darkness. Physiological processes that vary over the 24-hour day include activity, alertness, hormone secretion, organ physiology, and gene expression. These variations are not merely passive responses to a rhythmic environment, but instead reflect an underlying biological mechanism that can measure time in 24-hour increments. This mechanism orchestrates physiology to achieve predictive, rather than reactive, homeostasis.

Biological time measurement on the 24-hour time scale is achieved by a circadian timing mechanism. The term *circadian* is derived from the Latin words *circa* and *diem*, meaning "about a day," referring to the cycle length of these rhythms. Circadian rhythms are defined as rhythms that persist with a cycle length of approximately 24 hours in constant environmental conditions (Figs. 39.1 and 39.2). These rhythms are generated by a biological timing mechanism that is normally synchronized (entrained) to the 24-hour day by environmental cues. Light is the most widely used signal for entrainment of circadian clocks, but temperature, hormone levels, nutrient availability, and other cues can also affect oscillations.

In this chapter, we describe how circadian rhythms are generated, how they become synchronized to environmental cues, and their impact on physiology and behavior.

OVERVIEW OF THE MAMMALIAN CIRCADIAN TIMING SYSTEM

A biological timing system necessarily consists of an intrinsic clock mechanism that measures time, an input mechanism that allows the clock to become synchronized or reset by changes in the environment, and output pathways that lead to generation of overt rhythms such as daily changes in locomotor activity, sleep, and hormone levels. The mammalian circadian timing system is considerably more complicated than this simple linear scheme, because it is composed of a hierarchy of circadian oscillators (Fig. 39.3). At the pinnacle of this oscillatory system is a small brain area, the suprachiasmatic nuclei (SCN) of the anterior hypothalamus. The SCN are often called the master circadian clock, because these nuclei (one on each side of the brain) play a key role in coordinating oscillations in other tissues and in regulating behavior. Many other cells and tissues also have the capacity to display an approximately 24-hour rhythmicity. The molecular mechanism underlying these cell-autonomous circadian oscillations is a transcriptional feedback loop.

The primary input pathway to the SCN circadian clock is through retinal detection of light. Remarkably, the retinal photoreceptors that lead to visual image formation are not needed for circadian photoreception. Instead, a specialized population of retinal ganglion cells directly detect light, project to the SCN, and are necessary for photic entrainment of the SCN clock.

The main output of the SCN is encoded in neuronal firing rates. In addition, rhythmic neuropeptide secretion into the cerebrospinal fluid may be an important mechanism

© 2013 Elsevier Inc. All rights reserved.

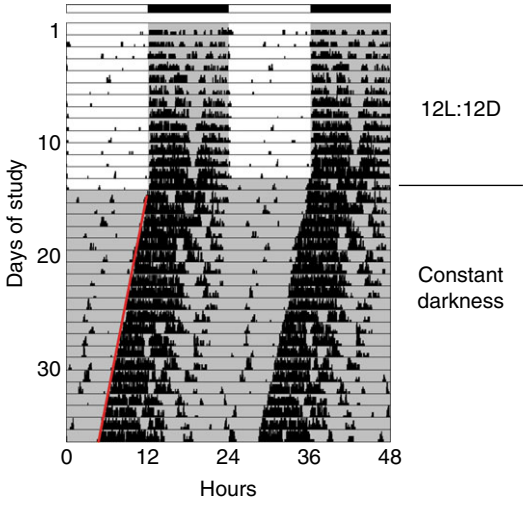


FIGURE 39.1 Locomotor activity rhythm of a mouse. This "double-plotted actogram" shows the activity pattern of an adult mouse housed individually in a cage with a running wheel. When the mouse rotates the wheel by running within it, a computerized system detects and records the wheel revolutions. Wheel revolutions are then plotted versus time, for weeks of data. Each horizontal line represents 48 hours of data. Wheel revolutions appear as bars coming up from the baseline, with their height proportional to the number of revolutions. On the first line of the record, the first 48 hours of data are shown. In the second line of the record, the data shown are from hours 24-72 of data collection. Thus, each 24-hour period of data after the first 24 hr is shown both to the right of the previous day's data and below it; this artificial reproduction of the data (double-plotting) helps to see the rhythmicity when the cycle length of rhythmicity differs from 24 hours. When housed in a 12 hours light, 12 hours darkness (12L:12D) lighting cycle, the mouse is active almost exclusively at night (shaded region). When the lighting cycle is disabled so the animal is in constant darkness (DD, starting on day 14), the animal continues to show rhythmicity but with a cycle length slightly less than 24 hours. This record reveals two defining principles of circadian rhythms: (1) circadian rhythmicity is intrinsic, rather than being generated in response to environmental variation, and (2) the circadian clock is normally synchronized to the 24-hour day length by light. It is easy to draw a "best-fit" line through the onset of activity each day (red line). Note the precision of the rhythmicity: the actual time of activity onset does not deviate far from this "best-fit" line, indicating the cycle length is measured with great accuracy. Spend a few minutes to think about how you would design a biological timekeeping system that can measure a day while having a variation between successive cycles that is less than 1% (15 minutes per 24 hours; the actual variation is often considerably less than this amount). *Figure courtesy of Jason DeBruyne, Morehouse School of Medicine.*

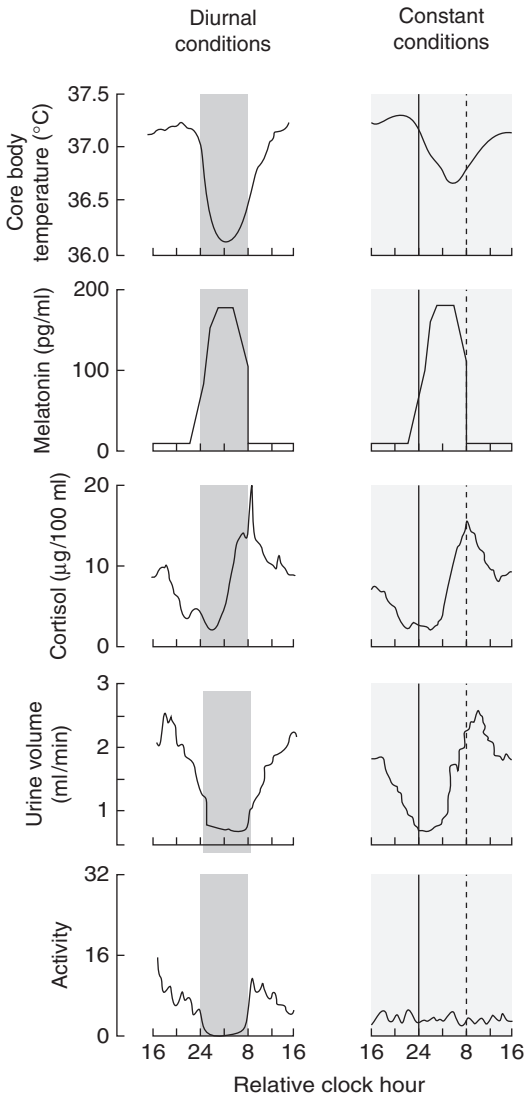


FIGURE 39.2 Diurnal and circadian rhythms in humans. The panels on the left show rhythms in several functions in human subjects maintained in a lighting cycle, with a period of sleep in the dark indicated by the shading. The panels on the right show the same endpoints studied in subjects maintained for 40 hours on a constant routine, with very low levels of illumination and with wakefulness maintained, while the subjects remain constantly in a semirecumbent posture and receive small meals at regular intervals throughout the study period. Persistence of the rhythms in constant conditions reveals they are true circadian rhythms. The apparent loss of the activity rhythm in constant conditions is due to the protocol-enforced inactivity of the subject. *Adapted from Czeisler, Buxton, & Khalsa, 2005, with permission Copyright 2005, Elsevier.*

for regulation of downstream targets. Anatomically, the outputs of the SCN are focused within the hypothalamus but have widespread influence, consistent with the pervasive influence of the SCN on physiological functions ranging from the modulation of cognitive function to neuroendocrine responses. Output rhythms dependent on the SCN synchronize autonomous oscillators in other tissues. Local circadian oscillators regulate gene expression in a tissue-specific manner. Microarray studies conducted on brain, liver, heart, and kidney tissue in mammals reveal that 5 to 10% of the genome is rhythmically transcribed, and many of the genes that are rhythmically expressed represent key, tissue-specific metabolic control points. Thus, the circadian timing system has a pervasive impact on physiological processes via its output pathways.

The Suprachiasmatic Nuclei are the Site of the Primary Circadian Pacemaker in Mammals

The SCN have been recognized as the master circadian clock in rodents since the early 1970s (see Weaver, 1998).

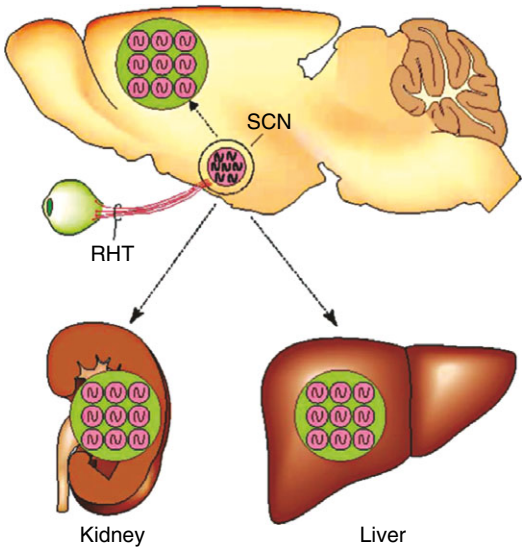


FIGURE 39.3 The mammalian circadian timing system consists of a hierarchy of oscillators. Oscillatory neurons in the SCN interact with each other to produce a set of coherent outputs. These outputs, which include behavioral and physiological rhythms, synchronize cell-autonomous oscillations in other brain regions and in peripheral tissues. *From Reppert & Weaver, 2002, with permission. Copyright 2002, Nature.*

At that time, "new" techniques for anterograde tract-tracing were used to assess retinal projections, and they revealed a direct retina-to-SCN projection, the retinohypothalamic tract. To determine if the SCN were an important node on the anatomical pathway for detection of light for regulating rhythms, Moore and Eichler (1972) destroyed the rat SCN and observed loss of rhythmic serum corticosterone in a population of rats. Using a similar approach, Stephan and Zucker (1972) discovered that the rhythm of drinking in individual rats was lost following destruction of the SCN. Together, these two studies demonstrated the necessity of the SCN for circadian rhythmicity. Hundreds of subsequent studies have confirmed the necessity of the SCN for rhythmicity in physiology and behavior (Weaver, 1998; Welsh, Takahashi, & Kay 2010).

SCN: Oscillator and Pacemaker

Lesion studies leading to loss of rhythmicity do not necessarily show that the SCN function as the primary circadian pacemaker. An alternative possibility is that the SCN could be a necessary element on a key output pathway leading to physiological rhythmicity. The presence of rhythms in SCN metabolic activity, electrical activity, and gene expression profiles *in vivo* do not distinguish between these possibilities. Studies demonstrating rhythms in neuropeptide secretion, metabolic activity, and electrical firing rate in SCN tissue maintained *in vitro* do show that the SCN contain a functional *oscillator*. The SCN do not simply oscillate, however: through its outputs, the SCN regulate rhythms in physiology and behavior, and thus serve as a circadian *pacemaker*. This unique pacemaker role of the SCN is revealed most clearly by studies showing that rodents made arrhythmic by SCN lesion can have rhythmicity restored by transplantation of fetal hypothalamic tissue containing the SCN into the third ventricle of the lesioned adult. Furthermore, transplants between hamsters with different circadian cycle lengths revealed that the period of restored rhythmicity is dictated by the genotype of the SCN tissue donor (Ralph, Foster, Davis, & Menaker 1990). Thus, the SCN serve as a pacemaker that communicates rhythmicity to tissues regulating behavioral rhythms.

Summary

The SCN function as the master circadian pacemaker in the mammalian brain. The SCN oscillate *in vivo* and also when placed *in vitro*. More importantly, however, the SCN generate output signals that lead to physiological and behavioral rhythms. The SCN is positioned at the interface between the outside world (detected by retinal photoreceptors) and light-insensitive effector tissues.

A HIERARCHY OF CELL-AUTONOMOUS CIRCADIAN OSCILLATORS

Single-Cell SCN Oscillators Are Coupled to Form a Coherent Clock

In rodents, the SCN contain approximately 10,000 neurons per side of the brain, with heterogeneity of neuropeptide phenotype, connectivity, and other properties (Silver & Schwartz, 2005; Morin, 2007). Despite this heterogeneity, most cells within the SCN are believed to have the capacity for cell-autonomous rhythmicity. Long-term rhythmicity in individual SCN neurons was first detected using extracellular recording of SCN neuronal firing rate (Fig. 39.4A). More recent studies reveal rhythmicity in single SCN cells using imaging methods designed to detect temporal variations in reporter gene expression (Fig. 39.4C).

The cycle length of individual oscillating SCN cells is more variable than the period of animal behavior or the period of rhythms detected from slices of SCN tissue maintained *in vitro*. Thus, some mechanisms for interaction among the heterogeneous SCN neurons must exist, allowing SCN neurons to synchronize their diverse periods.

Vasoactive Intestinal Peptide (VIP) plays a key role in synchronizing the activity of SCN neurons into a coherent, functional oscillator (Aton, Colwell, Harmar, Waschek, & Herzog, 2005; Maywood et al., 2006). VIP is expressed in a subset of neurons in the "core" region of the SCN (Fig. 39.5). VIP-containing neurons and other core neurons project heavily within the SCN core, to the surrounding "shell" region, and also beyond the SCN borders

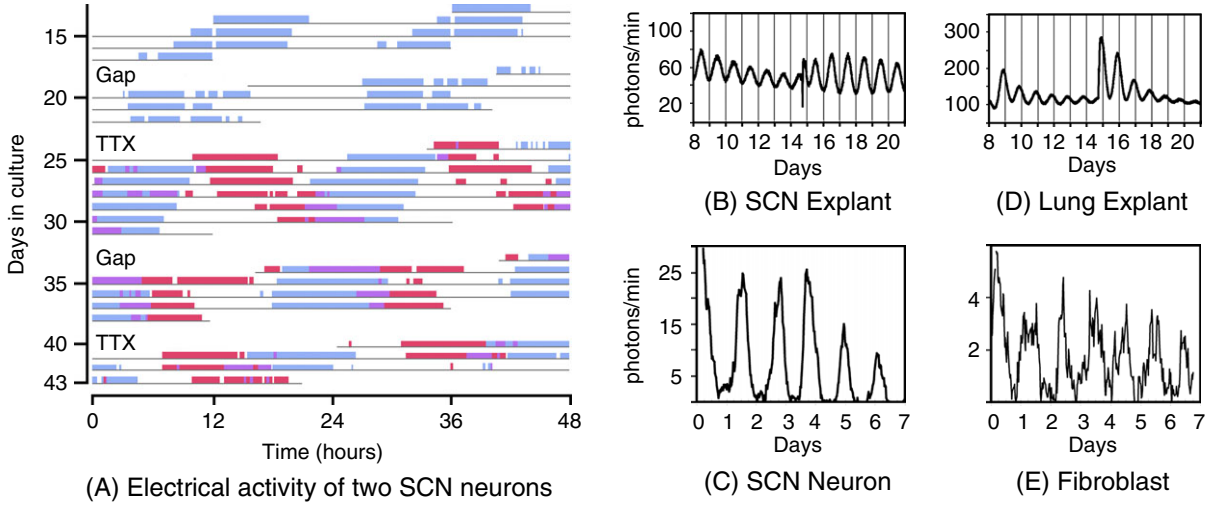


FIGURE 39.4 Cell-autonomous circadian rhythmicity *in vitro*. (A). Independently phased circadian rhythms in cultured SCN neurons. Spontaneous electrical activity of two SCN neurons in dispersed cell culture is shown in actogram-like format. Red bars represent one cell, while blue bars represent the other. The "activity" bars represent intervals when the neuronal firing rate for that cell is above its daily average. Where the active periods of the two cells overlap, they are shown in purple. Note that the two cells are cycling with independent period lengths. Gaps in the record represent days without recording. On two of these gaps, the culture was bathed with tetrodotoxin (TTX) for 2.5 days to block sodium-dependent action potentials. Note the molecular oscillations persist with unchanged period or phase despite TTX exposure. (*Adapted from D.K. Welsh, Logothetis, Meister, & Reppert, 1995, with permission. Copyright 1995, Elsevier.*) (B–E). Bioluminescence rhythms. Tissues and cells from mice with a rhythmically expressed luciferase reporter gene (encoding luciferase fused to PER2) were cultured in the presence of luciferin, and light emission was detected and plotted against time. Explants containing the SCN (B) or lung tissue (D) maintain rhythmicity in vitro. On the day labeled 14, culture media was changed, re-storing rhythm amplitude. Bioluminescence rhythms are also detectable, by imaging, from individual SCN neurons (C) and individual fibroblasts (E). *Adapted from A.C. Liu et al., 2007, with permission. Copyright 2007, Elsevier.*

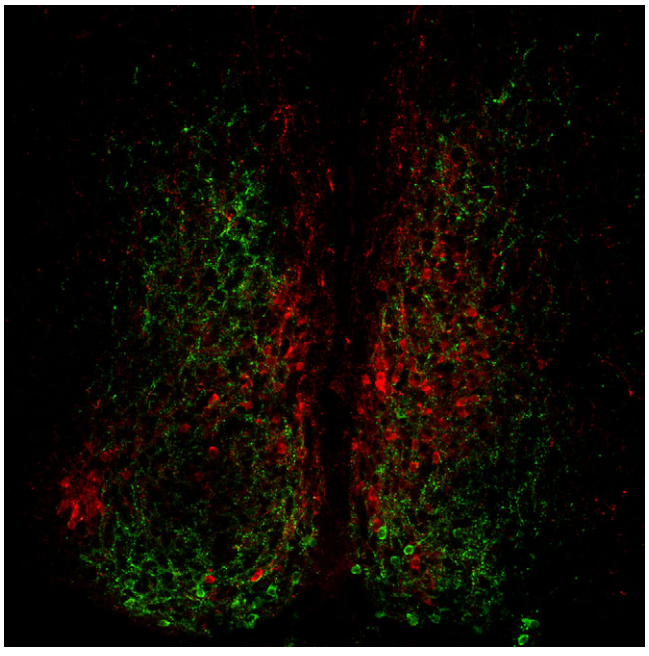


FIGURE 39.5 Neuropeptide expression in the mouse suprachiasmatic nucleus. Immunofluorescence image showing neurons expressing arginine vasopressin (AVP, red) in the dorsomedial aspect of the SCN, and neuronal cell bodies expressing VIP (green) in the ventral SCN. Note the extensive distribution of VIP-positive axons within the SCN. *Image courtesy of Nicola Smyllie and Mick Hastings, Laboratory of Molecular Biology, MRC, Cambridge, UK.*

(Fig. 39.5). Mice lacking VIP, or the VPAC2 receptor for VIP, have severely disrupted circadian rhythms. At the cellular level, the amplitude of individual neuronal oscillators is reduced in the absence of VIP or VPAC2 signaling, and the neurons are not coordinated to one another (Aton & Herzog 2005; Maywood et al., 2006). These findings strongly suggest that a "coupling deficit" leads to altered circadian rhythmicity in these lines of mice. VIP/VPAC2 acts through elevation of cAMP levels in SCN (Atkinson et al., 2011). Interestingly, the receptor for the insect neuropeptide Pigment Dispersing Factor (PDF) is closely related in structure to the VIP receptor (Helfrich-Förster, 2005). Moreover, PDF plays a role in the *Drosophila* (fruit fly) circadian system, which is strikingly similar to the role played by VIP in the mammalian SCN (Box 39.1). This indicates that the mechanisms underlying the communication between circadian neurons were established very early during the evolution of animals, at least 600 million years ago.

The entire population of SCN neurons usually functions in concert as an orchestrating oscillator. Under specific conditions, however, SCN cells can become coupled into two functionally distinct subpopulations that can independently influence circadian behavior. For example, hamsters exposed to constant light for long periods of time display "splitting," leading to two peaks of locomotor activity per 24 hours. Molecular analysis has revealed that

BOX 39.1

NEURONAL CONTROL OF *DROSOPHILA* CIRCADIAN BEHAVIOR

One of the greatest strengths of *Drosophila* as a model to study circadian behavior (and more generally any behavior) is the remarkably sophisticated set of tools available to manipulate specific populations of neurons. Using tissue-specific "drivers" to express various genes, neurons can be eliminated, electrically silenced or forced to fire action potentials, for example. Levels of gene expression can also be manipulated by overexpression or RNA interference. With this wide panoply of genetic tools, researchers are beginning to understand the role of the complex mosaic of circadian neurons.

There are about 150 circadian neurons in the fly brain, divided in several groups named by their anatomical locations (see Figure B39.1). These clusters of circadian neurons can be further divided based on their morphology, and the neuropeptides or neurotransmitters they express. The small ventral Lateral Neurons (s-LN_vs) function as pacemaker neurons in constant darkness (DD): they are necessary and sufficient for rhythmic locomotor behavior to persist in DD. These neurons secrete the neuropeptide Pigment Dispersing Factor (PDF). PDF is necessary for self-sustained behavioral rhythms and to synchronize molecular circadian rhythms in the brain. Interestingly, the phenotypes of flies mutant for either PDF or its receptor (PDFR, which is expressed on most circadian neurons) are remarkably similar to those of mice missing VIP or its receptor (VIPR): arrhythmicity and desynchronization between circadian neurons. Moreover, VIPR and PDFR share extensive sequences homologies. Thus, functional and molecular conservation of circadian rhythm mechanisms between mammals and insects is not limited to the circadian molecular pacemaker (see Box 39.2), but extend to circadian neural circuits.

Why have so many different circadian neurons if just 8 of them (the s-LN_vs) can generate robust behavioral rhythms in DD? A first answer is the need to divide tasks. Fruit flies are mostly active at dawn and dusk under Light:Dark (LD) cycles. The PDF positive s-LN_vs drive morning activity, but it is a different group of cells that drive evening activity: the dorsal Lateral Neurons (LN_ds). The s-LN_vs and the LN_ds might thus increase their electrical activity at different times of the day to generate morning and evening bouts of activity.

A second reason for the need of a complex circadian neural network is adaptation to an ever-changing environment. Day length (photoperiod) and overall temperatures cycle during the course of the year at most latitudes, and random weather changes influence daily environmental conditions. It is becoming increasingly clear that one of the main purposes for the diversity of circadian neurons is the integration of environmental inputs and a proper adaptation to external conditions. For example, some circadian neurons (DN2s, LPNs) are particularly sensitive to temperature cycles. Others (l-LN_vs, DN_1s) appear to be important for responses to light inputs. Finally, the contribution of a subset of DN_1s to circadian locomotor behavior is dependent on both ambient light intensity and temperature. Interestingly, the hierarchy between circadian neurons can change in response to light exposure. In the dark, or under short photoperiod, the PDF positive s-LN_vs play a particularly important role in the control of circadian behavior, while under constant light or under long photoperiod PDF negative circadian neurons (LN_ds, DN_1s) are dominant. This presumably helps flies to cope with seasonal changes in day length. Whether similar mechanisms act in mammals is not yet clear, but specific populations of SCN neurons adjust their firing cycles in response to changes in photoperiods.

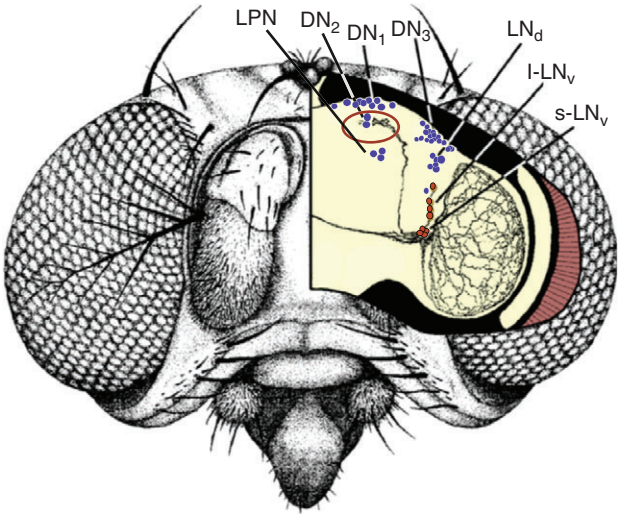


FIGURE B39.1 **Circadian neurons in the brain of *Drosophila melanogaster*.** Neurons expressing circadian proteins are shown in blue (Dorsal Neurons 1–3 [DN_1, DN_2, DN_3], dorsal Lateral Neurons [LNd], and Lateral Posterior Neurons [LPN]). Neurons expressing circadian proteins and the circadian neuropeptide Pigment Dispersing Factor (PDF) are shown in red (small and large ventral Lateral Neurons [s-LN_v and l-LN_v]). The PDF-positive neural projections are drawn in black. The most important projections for the synchronization of the circadian network are the s-LN_v projections directed toward the DN_1 and DN_2 groups. Indeed, PDF level oscillates in the termini of these projections (circled in red), presumably because PDF is rhythmically released. *Adapted from C. Helfrich-Förster, 2005, with permission. Copyright 2005, Elsevier.*

Patrick Emery

Further Readings

Nitabach, M. N. & Taghert, P. H. (2008). Organization of the *Drosophila* circadian control circuit. *Current Biology, 18,* R84–93.

Zhang, Y. & Emery, P. (2012). Molecular and neural control of insect circadian rhythms. In L. I. Gilbert (Ed.), *Insect Molecular Biology and Biochemistry* (pp. 513–551). San Diego, Academic Press.

the left and right SCN are oscillating in opposite phases in "split" hamsters, with each side regulating locomotor activity (de la Iglesia, Meyer, Carpino, & Schwartz, 2000). In rats, complex lighting cycles can be used to dissociate ventrolateral and dorsomedial subdivisions of the SCN from each other, resulting in simultaneous expression of rhythms with two different periods; each activity component is apparently driven by a different region of the SCN. Imaging studies reveal waves of gene expression progressing through the SCN *in vitro*, and subpopulations of neurons are coupled with differing relationships to the environmental lighting cycle. These experiments emphasize that the multi-oscillatory nature of the circadian system can be appreciated even within the SCN, and that coupling among SCN neurons normally prevents conflicting output signals.

Peripheral Oscillators

Despite sporadic reports of rhythmicity in isolated mammalian tissues, for years the dogma in the circadian field was that only the SCN possessed the capacity for self-sustaining, autonomous rhythmicity. The identification of a family of mammalian *Period* genes helped to change this perception. The mammalian *Period* genes are widely expressed (Zylka, Shearman, Weaver, & Reppert, 1998), suggesting that rhythmicity could occur in additional tissues. In addition, widespread, cell-automonous circadian clocks were demonstrated in *Drosophila* at about the same time (Plautz, Kaneko, Hall, & Kay, 1997).

That the capacity for circadian rhythmicity is not limited to the SCN in mammals was clearly demonstrated in a landmark study by Balsalobre, Damiola, and Schibler (1998). The field was shocked to learn that brief exposure to a high concentration of serum ("serum shock") could lead to rhythmic gene expression in cultured fibroblasts. More recent single-cell analysis revealed that fibroblasts have ongoing (but uncoordinated) rhythms, and a serum pulse merely synchronizes these rhythms so that rhythmicity can be detected in the population (Nagoshi et al., 2004; Welsh et al., 2004; see Fig. 39.4). The serum shock is not a unique stimulus; acute activation of each of several pathways can induce rhythmicity in fibroblasts, suggesting that multiple signaling pathways converge upon the oscillator.

It is now generally believed that most mammalian cell types are capable of circadian oscillation. In normal circumstances, the rhythmicity in tissues outside the SCN is synchronized by SCN-dependent output signals. These signals include physiological and behavioral rhythms (e.g., body temperature and food intake) and daily fluctuations in hormone levels (e.g., glucocorticoids and melatonin). Thus, rhythms controlled by the SCN synchronize molecular rhythmicity among cells within a peripheral tissue (Fig. 39.3). In the absence of rhythmic input from the SCN, oscillators in some tissues become desynchronized, and thus the organ as a whole loses detectable rhythmicity. In the olfactory bulb, however, interactions among oscillating cells are sufficiently robust to maintain coordinated rhythmicity (Granados-Fuentes, Prolo, Abraham, & Herzog, 2004).

While the SCN function as the master circadian pacemaker in most circumstances, there is now compelling evidence that oscillators outside the SCN can regulate rhythmic behavior (Stephan, 2002; Taratoglu, Davidson, Benvenuto, & Menaker, 2006). While these oscillatory mechanisms are overshadowed in the presence of a functioning SCN, these studies reveal the potential impact of tissue-autonomous oscillators outside the SCN. Our understanding of the implications of the hierarchical nature of the mammalian circadian timing system is advancing rapidly.

Coordination of rhythms within a tissue may optimize organ-level physiological processes, while a circadian timing system in general may achieve "internal temporal order" among organs so that organs are prepared to function most efficiently. Circadian misalignment and disruption of internal coordination occur with shift work and following transmeridian travel (jet lag) (see Box 39.3).

Summary

Many cell types possess the capacity for oscillation with a cycle length near 24 hours. Neurons within the SCN oscillate and also communicate with one another, such that the SCN are able to provide a coherent output resulting in rhythmicity of other functions. Oscillating cells in other tissues generally require SCN-dependent signals, in the form of physiological rhythms, to remain synchronized, both within and between tissues.

THE MOLECULAR BASIS FOR CIRCADIAN OSCILLATION IS A TRANSCRIPTIONAL FEEDBACK LOOP

The preceding discussion emphasizes that circadian rhythmicity occurs within single, isolated cells. But how do individual cells measure 24-hour time intervals? Studies conducted in many labs over the past 15 years have revealed a great deal about the molecular mechanisms underlying mammalian circadian rhythmicity (Fig. 39.6).

Progress made in understanding the circadian clock mechanisms in the fruit fly, *Drosophila melanogaster*, was instrumental in developing our current understanding of mammalian clock mechanisms (Box 39.2). Most of the genes involved in the fly circadian mechanism have mammalian homologs, and their roles in circadian function have now been examined in mammals. While there have been some alterations in gene products performing specific functions, and gene duplication in mammals has led

to greater complexity, the general features of the core clock mechanisms are similar between flies and mammals (Box 39.2). This demonstrates again the ancient origin of circadian timing mechanisms in the animal kingdom.

The molecular mechanism for measuring time on a 24-hour time-scale is based on a self-sustaining transcriptional/translational feedback loop present within individual cells (Fig. 39.6). According to the current model for this molecular feedback loop (Reppert & Weaver, 2002; DeBruyne, 2008; Welsh et al., 2010), pairs of basic helix-loop-helix, PAS-domain containing (bHLH-PAS) transcription factors dimerize and activate transcription of several genes, including *Period* and *Cryptochrome* genes, through specific promoter elements called E-boxes. CLOCK and BMAL1 are the major transcriptional activator complex, but in a subset of SCN neurons, NPAS2 and BMAL1 heterodimers form and can maintain rhythmicity. High-amplitude rhythmicity in the expression of two *Period* genes (*Per1* and *Per2*; Fig. 39.7) is believed to be rate-limiting for formation of complexes between the PERIOD (PER) and CRYPTOCHROME (CRY) proteins. PER:CRY-containing protein complexes recruit other proteins, and this complex disrupts the activity of the CLOCK:BMAL1 complex. The temporally regulated nuclear accumulation of PER:CRY complexes is thought to be critical for closing the feedback loop. Thus, molecular oscillations occur with a cycle length of approximately 24 hours, in which a phase of CLOCK/NPAS2:BMAL1-mediated transcriptional activation is followed by phase of PER:CRY-mediated repression. It is important to note that this transcriptional feedback loop regulates additional genes, including other transcription factors, thus generating rhythmic expression of a large number of genes (Jin et al., 1999; Ueda et al., 2005; Kornmann, Schaad, Bujard, Takahashi, & Schibler, 2007; Miller et al., 2007; Atwood et al., 2011).

Post-translational modifications build a time delay into the circadian oscillation. Without this delay, transcription of *Per* genes would rapidly lead to nuclear accumulation of PER:CRY complexes and a premature inhibition of transcription. In this case, the clock would either cycle with a very short period or the system would equilibrate to a nonrhythmic steady state. Phosphorylation of the PER proteins is critical for this time delay. Phosphorylation alters their intracellular localization, protein:protein interactions, and stability, and thereby contributes to regulating circadian cycle length. Rhythmic phosphorylation of PER proteins is due at least in part to the activity of casein kinase 1 delta (CK1δ) and CK1 epsilon (CK1ε) and mutations of these kinase genes or their pharmacological inhibition alters circadian cycle length in flies, hamsters, mice, and humans. A circadian-based sleep disorder in humans is due to alterations in CK1-mediated phosphorylation of PER2 proteins, leading to rhythms having an unusually early phase (Box 39.3). Phosphorylation of clock proteins alters their activity and can also target them for proteosomal degradation; mutations in the gene encoding FBXL3, which participates in phosphorylation-dependent ubiquitination of CRY proteins, slow the circadian cycle in mice (Gatfield & Schibler, 2007). β-TRCP is responsible for the ubiquitination and degradation of PER proteins. The fine balance between kinases (including CK1δ/ε and CK2) and phosphatases (notably protein phosphatase 1), and subsequent phosphorylation-dependent degradation contribute greatly to regulating circadian period length.

An interlocked, second feedback loop is important for the rhythmic regulation of BMAL1. Two members of the orphan nuclear receptor gene family, *Rev-erb-alpha* and *Rora,* are among the genes regulated by CLOCK and BMAL1. The protein products of these genes compete for binding to an orphan nuclear receptor binding site in the promoter of the *Bmal1* gene and have antagonistic effects on *Bmal1* gene expression. Overexpression of the repressor, REV-ERB-α in the liver greatly reduces *Bmal1* expression, leading to liver-specific loss of rhythmicity (Kornmann et al., 2007). While BMAL1 is necessary for rhythmicity, *rhythmic* BMAL1 production appears relatively unimportant for maintaining rhythmicity.

Molecular Redundancies Revealed by Studies of Knockout Mice

The importance of the genes discussed above in the circadian mechanism has been revealed in part through studies of mice with targeted gene disruption. These studies reveal that, in most cases, there is at least partial redundancy of function among members of these gene families (Table 39.1). Mice homozygous for mutation of either *mPer1* or *mPer2* alone have period defects and/or a gradual loss of rhythmicity, while mice homozygous for disruption of both *mPer1* and *mPer2* become arrhythmic immediately in constant darkness. Notably, however, the phenotype of mice homozygous for mutation of either *Per* gene is markedly affected by genetic background; single *Per* mutations are almost without effect on the C57BL/6J background (Pendergast, Friday, & Yamazaki, 2010), indicating redundancy between the PER proteins.

Mice with targeted disruption of either *mCry1* or *mCry2* have altered circadian period lengths (~22.5 and ~24.5 hours, respectively), but the double-mutant mice completely lose circadian rhythmicity immediately upon placement in constant darkness.

While mice homozygous for a semidominant mutation in the *Clock* gene ($Clock^{\Delta 19}$) have significant behavioral and molecular phenotypes, mice completely lacking CLOCK protein retain behavioral rhythmicity (Fig. 39.8). As noted earlier, the closely related bHLH-PAS transcription factor NPAS2 (also called MOP4) maintains circadian rhythms in the SCN, even in the absence of CLOCK (Fig. 39.8).

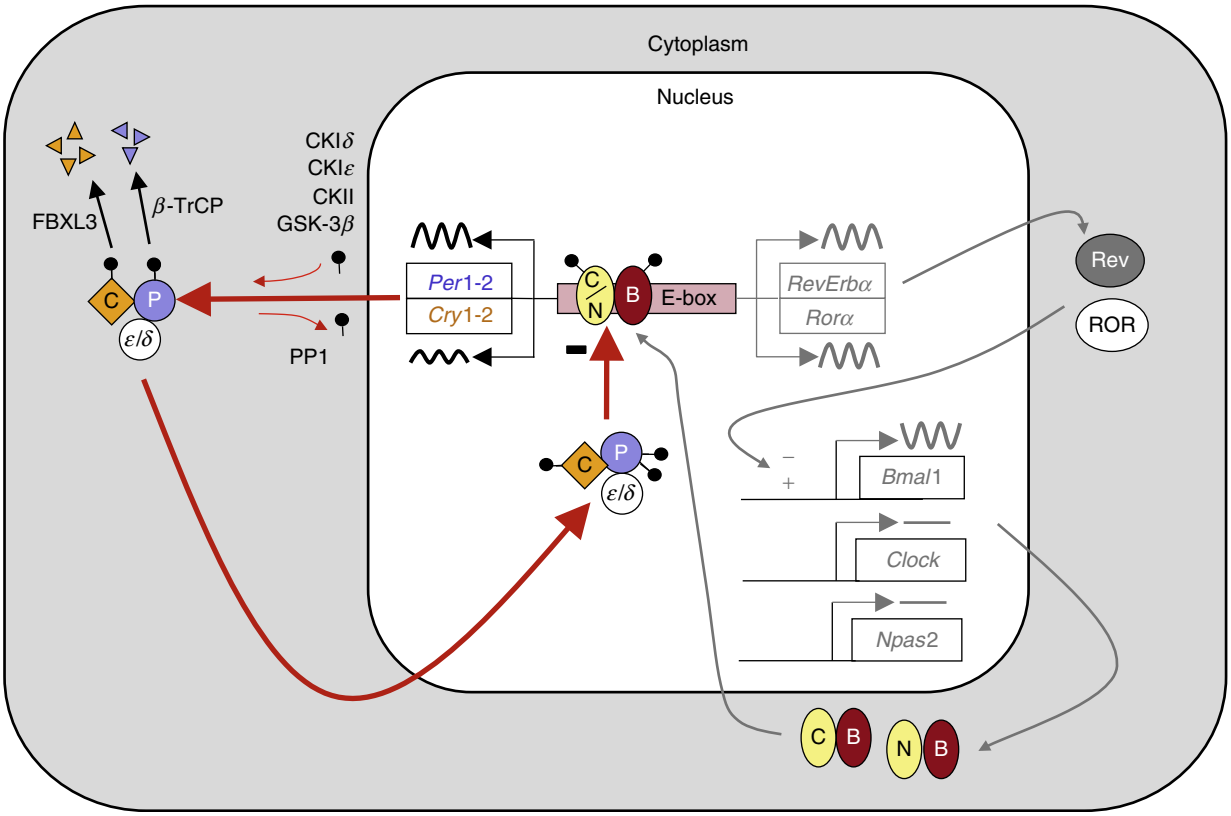


FIGURE 39.6 The molecular mechanism for circadian rhythms in mammals is a transcriptional feedback loop. Transcriptional activation leads to expression of *Per* and *Cry* genes. Posttranslational modifications, including phosphorylation (black lollipop) mediated in part by casein kinases (CKIδ, CKIε, CKII) and GSK-3β, affect the interactions of PER (P in blue circle) and CRY (C in gold diamond) proteins, promote their proteosomal degradation by F-box proteins (β-TrCP and FBXL3), and regulate nuclear entry of the inhibitory PER:CRY complex. The high-amplitude rhythm of PER production controls the timing of negative feedback. Collectively, these events lead to a delayed, negative feedback to shut off the transcriptional activation, resulting in a negative feedback loop (red arrows). The positive drive to the system comes from the transcription factors CLOCK (C in yellow oval) and BMAL1 (B in red oval), or NPAS2 (N in yellow oval) and BMAL1. The orphan nuclear receptors RORA (ROR) and REVERB-alpha (Rev) control the rhythmic expression of *Bmal1*, with peak levels occurring opposite to the peak in *Per* expression. The antiphase rhythmicity of this second feedback loop is not essential for rhythm generation, but BMAL1 itself is necessary.

BOX 39.2

MOLECULAR MECHANISMS OF CIRCADIAN TIMEKEEPING IN *DROSOPHILA*

Virtually nothing was known about the genetic and molecular mechanisms controlling circadian rhythms until the late 1960s, when Seymour Benzer and his graduate student, Ron Konopka, identified the first mutations affecting circadian behavior. Most scientists at the time were skeptical that individual genes could affect complex behaviors. Benzer was nevertheless determined to identify the genetic underpinnings of behavior, and embarked on an ambitious program using the animal model of choice for genetic studies, the fruit fly *Drosophila melanogaster*. Benzer and colleagues exposed large numbers of flies to a chemical mutagen to isolate mutant lines with abnormal behavior in various assays. To identify circadian genes, Konopka screened for mutant fly lines that had lost the ability to predict dawn, as assessed by observing the time of day at which the young flies emerged from their pupal cases. Using this approach, Konopka and Benzer isolated three mutant lines with abnormal circadian rhythms. In one line, the flies emerged from their pupal cases too early, in a second line they were too late, while in a third line the flies emerged at random times. All three mutations mapped to the same genetic locus, named *period (per)*. Importantly, the three *period*

BOX 39.2 (*cont'd*)

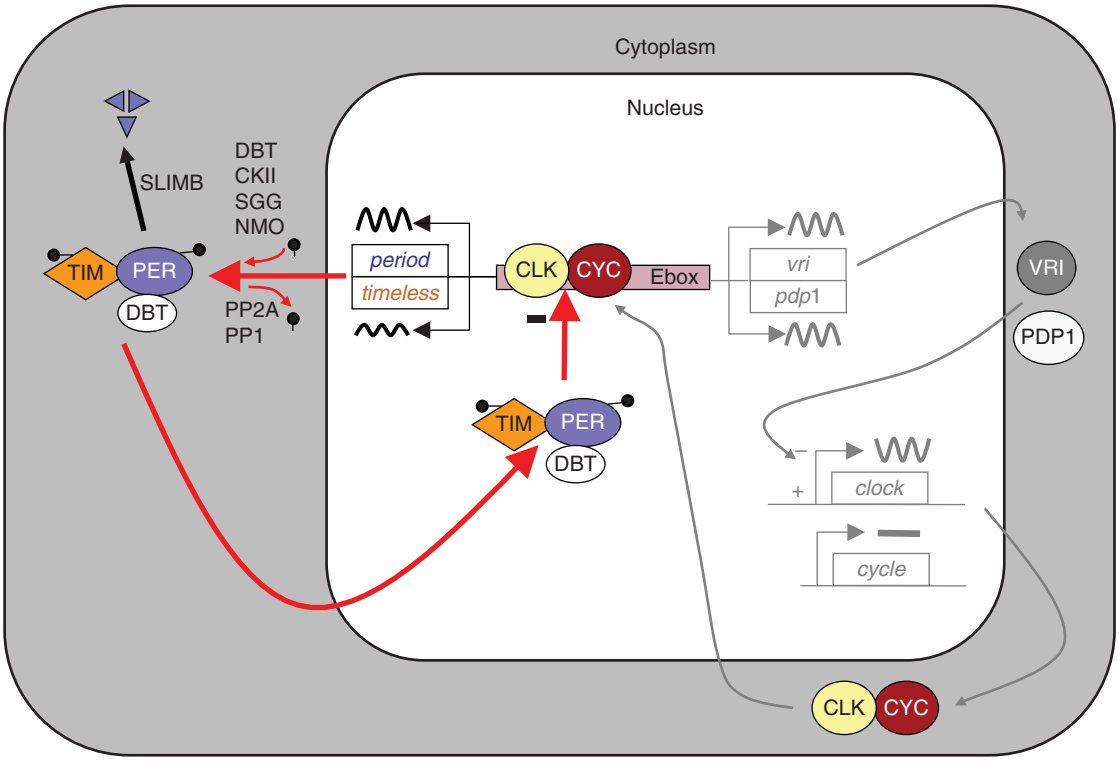


FIGURE B39.2 **Model of the molecular circadian clock in *Drosophila*.** In the core transcriptional feedback loop (left side of the figure), CLOCK (CLK) and CYCLE (CYC) positively regulate *period* (*per*) and *timeless* (*tim*) transcription. PER is phosphorylated by DOUBLETIME (DBT), CASEIN KINASE II (CKII]) and NEMO (NMO), while TIM is phosphorylated by SHAGGY (SGG) and CKII. PER and TIM phosphorylation levels are negatively regulated by PROTEIN PHOSPHATASE 1 and 2A (PP1 and PP2A). Hyperphosphorylated PER is ubiquitinated by SLIMB and degraded by the proteasome. PER and TIM enter into the nucleus with DBT to inhibit the activity of CLK/CYC. In a secondary feedback loop shown in grey, VRILLE (VRI) and PAR DOMAIN PROTEIN 1ε (PDP1) regulate *clk* transcription.

mutants had similar circadian defects in their locomotor activity rhythms, which provided strong evidence that Konopka and Benzer had identified a gene central to the circadian pacemaker.

Benzer's pioneering idea of using unbiased genetic screens to identify genes involved in circadian behavior has led to identification of additional "clock genes" by several other labs. The subsequent genetic, molecular, and biochemical studies have resulted in a remarkably detailed understanding of the *Drosophila* circadian pacemaker (see Figure B39.2). CLOCK (CLK) and CYCLE (CYC) are two basic helix-loop-helix transcription factors that heterodimerize and bind to E-boxes within the *per* and *timeless* (*tim*) promoters. PER and TIM dimerize in the cytoplasm, with TIM stabilizing PER. Several kinases (DOUBLETIME, CASEIN KINASE 2, NEMO, and SHAGGY [the *Drosophila* homolog of glycogen synthase kinase 3]) and at least two phosphatases (PROTEIN PHOSPHATASE 2A and PROTEIN PHOSPHATASE 1) regulate the stability and thus the rate of PER and TIM accumulation, and the timing of their nuclear entry. Once in the nucleus, the PER/TIM heterodimer binds to the CLK/CYC dimer and disrupts its transcriptional activity and its interaction with E-box sequences. The inhibition of *per* and *tim* gene expression is maintained for several hours until the hyperphosphorylated and unstable PER and TIM proteins decay. CLK/CYC can once again bind to E-box sequences and restart the circadian cycle. The kinase DOUBLETIME, which translocates into the nucleus with PER and TIM, also plays an important role for repression by regulating CLK phosphorylation and degradation. The CLK/CYC, PER/TIM feedback loop is essential for maintaining rhythms in *Drosophila*. A secondary feedback loop might contribute to the precision and robustness of circadian oscillations through the regulation of *clk* mRNA levels. The *Drosophila* molecular clock is primarily synchronized to environmental light/dark cycles by an intracellular photoreceptor, CRYPTOCHROME (CRY). Light activates CRY, and CRY then binds to and triggers the rapid degradation of TIM. Since TIM stabilizes PER, the degradation of TIM also results in a reduction in PER levels that resets the circadian pacemaker.

BOX 39.2 (*cont'd*)

As the first components of the mammalian molecular clock were identified in the late 1990s, it quickly became clear that important concepts learned from the *Drosophila* circadian clock also apply to mammals. The circadian pacemaker of both *Drosophila* and mammals involve homologous proteins in analogous transcriptional feedback loops. Strikingly, the first identified genetic cause of a human circadian rhythm disorder, Familial Advanced Sleep Phase Syndrome, is a mutation in human *Per2*, a homolog of the *Drosophila per* gene.

Ania Busza and Patrick Emery

Further Readings

Konopka, R. J., & Benzer, S. (1971). Clock mutants of *Drosophila melanogaster*, *Proc Natl Acad Sci USA 68*, 2112–2116.

Hardin, P. E. (2005). The circadian timekeeping system of *Drosophila*. *Current Biology*, *15*, R714–722.

Zhang, Y., & Emery, P. (2011). Molecular and Neural Control of Insect Circadian Rhythms. In L. I. Gilbert (Ed.), *Insect molecular biology and biochemistry* (pp. 513–551). Academic Press.

Weiner, J. (2000). *Time, love, memory: a great biologist and his quest for the origins of behavior*. New York: Alfred A. Knopf.

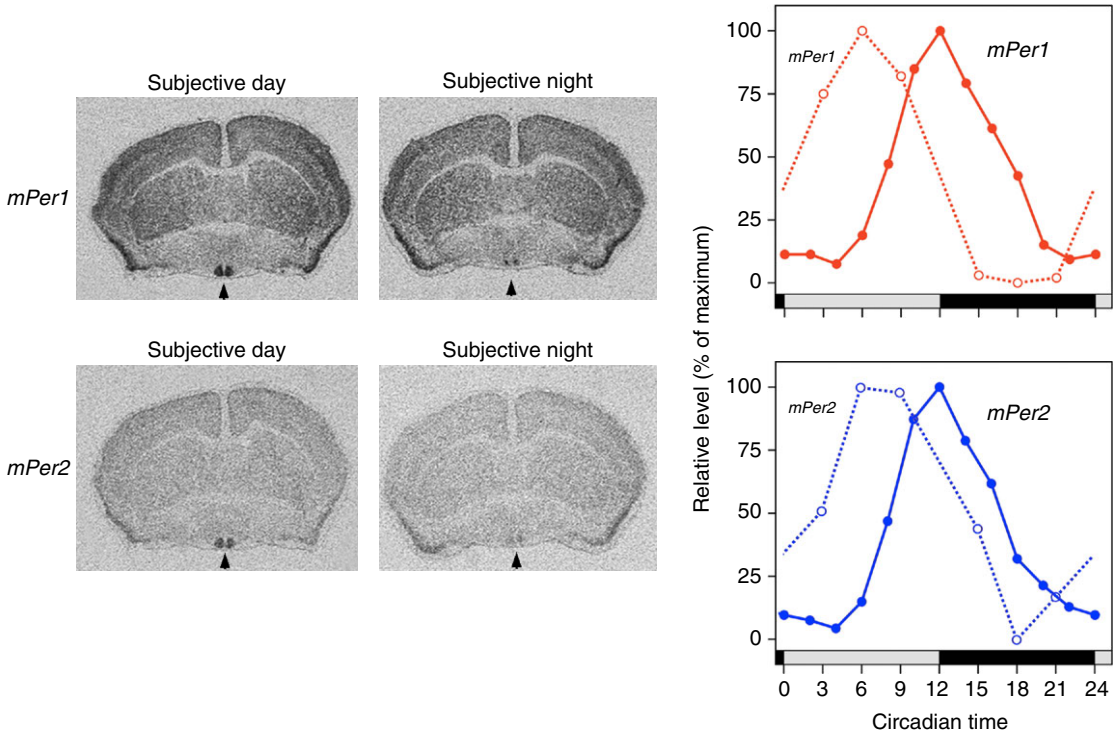


FIGURE 39.7 High-amplitude circadian rhythms of *mPer* gene expression in the mouse SCN. **Left:** Adjacent sections through the SCN from brains collected at the time of peak Per transcript levels during the subjective day (left column) or during the middle of the subjective night (right column) were processed for *in situ* hybridization to detect *mPer1* and *mPer2* mRNAs. "Subjective" refers to the fact that the animals were in constant darkness on the day of tissue collection, rather than in a lighting cycle. The SCN are indicated by an arrowhead in each panel. *Adapted from Shearman, Zylka, Weaver, Kolakowski, & Reppert, 1997, with permission. Copyright 1997, Elsevier.* **Right:** *Per* RNA (open circles and dashed lines) and PER protein (filled circles and solid lines) rhythms in mouse SCN. *mPer1* (upper panel) and *mPer2* (lower panel) RNA levels were assessed by *in situ* hybridization from samples collected on the first day in constant darkness. Protein levels were determined by counting nuclei within the SCN from sections processed for immunohistochemical detection of mPER1 and mPER2. The bar below the panel represents the lighting cycle the animals were exposed to on days prior to the day of tissue collection. Circadian Time 0–12 corresponds to subjective day, when the lights would have been on had the animals remained in a lighting cycle. (*Data from Shearman et al., 1997 and Hastings, Field, Maywood, Weaver, & Reppert, 1999.*)

BOX 39.3

CIRCADIAN-BASED SLEEP DISORDERS, CLOCK GENES AND NEUROPSYCHIATRIC DISEASES

The timing of sleep observed in adult humans normally is from approximately midnight to 8 a.m. Most people will voluntarily move sleep to a later time when free of work constraints and readily readjust their schedule back to an earlier time when the workweek begins again. There are developmental changes in the "normal" timing of sleep, with adolescents being prone to a later phase and the elderly typically assuming an advanced (earlier) phase. These developmental changes in preferred phase of activity and sleep mirror the natural variation within adult populations, with early-rising "larks" and late-rising "night owls" being readily recognizable patterns. At the extreme, however, these differences in timing of activity and sleep can disrupt social life and employment. These extreme cases are referred to as circadian-based sleep disorders (see Figure B39.3).

Advanced sleep phase syndrome (ASPS) is characterized by early-morning awakening and an inability to maintain wakefulness into the evening (the typical wake interval is from 0400 to 2000; see top panel of figure). Studies of familial cases of ASPS have identified two genes that, when mutated, contribute to ASPS. Remarkably, these genes were already discussed in the context of the core circadian feedback loop in mice. The mutations defined in humans that lead to familial ASPS are a Serine to Glycine mutation at position 662 of human PER2 (within the casein kinase interacting domain of PER2), and a Threonine to Alanine mutation at position 44 of casein kinase I delta (a mutation that increases the ability of the kinase to phosphorylate PER2).

Delayed sleep phase syndrome (DSPS) is characterized by late awakening and late bedtimes (typical wake interval from noon to 0400), and an apparent inability to reset the clock to earlier times of day. This disorder frequently begins in adolescence or young adulthood and tends to be sporadic rather than familial.

"Non-24-hour sleep wake syndrome" has also been described, in which individuals fail to entrain to the 24-hour day (see bottom panel of figure). This condition is relatively frequent in blind individuals. Affected individuals have intervals where their biological day and night coincide with the solar day and night, but then their free-running period (usually ~24.5 hours) causes them to drift out of phase with the environment. This results in periods of insomnia and daytime drowsiness (when the individuals are out-of-phase) alternating with periods of coordination with the environment. Melatonin, administered at 24-hour intervals, is useful in synchronizing some blind individuals to a 24-hour cycle length.

Several studies have identified genetic differences between "owls" and "larks." Specific haplotypes of the *Per3* gene, in particular, have been associated with delayed diurnal preference and with delayed sleep phase syndrome. Thus, even within the normal range of sleep timing, phase preference may have a significant genetic component.

Circadian rhythms are phase-advanced in depression, and rhythm alterations occur in other neuropsychiatric and neurodegenerative diseases. Recent studies reveal increased prevalence of certain disorders in subjects with specific alleles of circadian-relevant diseases. Circadian rhythm alterations may be a risk factor for neuropsychiatric disease.

David R. Weaver

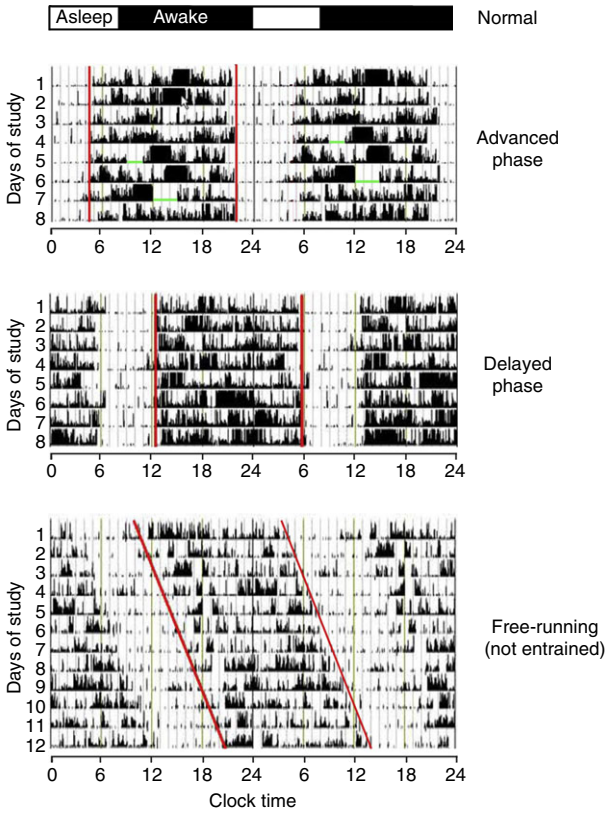


FIGURE B39.3 Disorders of sleep timing in humans. The typical timing of sleep and wake is shown at the top of the figure. Below are double-plotted activity records of three individuals. Activity was recorded by a motion-sensitive wristwatch. Activity appears as vertical deflections in the records. Intervals where the watch was taken off are indicated by green bars. Red lines mark activity onset and offset. The top actogram illustrates an individual with advanced sleep phase syndrome; activity occurs between 0400 and 2000. The second pattern shows delayed sleep phase syndrome; activity occurs from 1300 to 0600. The lower actogram shows an individual that is free-running due to failure to entrain to environmental cues. *Modified from Sleep, Medicine, vol 8, Barion, A., & Zee, P.C., "A clinical approach to circadian rhythm sleep disorders," pp. 566–677, Copyright 2007, with permission from Elsevier.*

BOX 39.3 *(cont'd)*

Further Reading

Dijk, D-J, & Archer, S. N. (2010). PERIOD3, circadian phenotypes, and sleep homeostasis. *Sleep Medicine Reviews, 14,* 151–160.

Jones, C. R., Campbell, S. S., Zone, S. E., Cooper, F., DeSano, A., Murphy, P. J., … Ptacek, L. J. (1999). Familial advanced sleep-phase syndrome: A short-period circadian rhythm variant in humans. *Nature Medicine, 5,* 1062–1065.

Kennaway, D. J. (2010). Clock genes at the heart of depression. *Journal of Psychopharmacology, 24*(2 Suppl), 5–14.

Toh, K. L., Jones, C. R., He, Y., Eide, E. J., Hinz, W. A., Virshup, D. M., … Fu, Y. H. (2001). An *hPer2* phosphorylation site mutation in familial advanced sleep phase syndrome. *Science, 291,* 1040–1043.

Xu, Y., Padiath, Q. S., Shapiro, R. E., Jones, C. R., Wu, S. C., Saigoh, N., … Fu, Y. H. (2005). Functional consequences of a CKIδ mutation causing familial advanced sleep phase syndrome. *Nature, 434,* 640–644.

Menet, J. S., & Rosbash, M. (2011). When brain clocks lose track of time: cause or consequence of neuropsychiatric disorders. *Current Opinion in Neurobiology, 21,* 1–9.

CLOCK appears to be much more abundant than NPAS2 in the SCN, and transcriptional regulation occurs primarily through CLOCK:BMAL1 heterodimers in SCN and other tissues. BMAL1 is necessary for maintenance of behavioral rhythmicity (Bunger et al., 2000). Overexpression of BMAL2, a protein closely related to BMAL1, can restore rhythmicity in BMAL1-deficient mice (Shi et al., 2010). BMAL2 expression is greatly reduced in BMAL1-deficient tissues; the BMAL1-deficient mouse may be a functional double-knockout, explaining this apparent departure from the trend for functional redundancy among gene family members.

Posttranslational modifications were discussed briefly above. Disruption of CK1δ cause period lengthening,

TABLE 39.1 Gene Families Regulating Mammalian Circadian Rhythms

Gene Family	Gene	Mutation	Allele Type	Effect on Behavioral Rhythmicity in DD
bHLH/PAS	*Clock*	Clock$^{\Delta 19}$(exon skipped)	Dominant-negative	Long period, then arrhythmic
	Clock	Targeted disruption	Null	Slightly short period
	Npas2 (*Mop4*)	Targeted disruption	Deletion	Slightly short period
	Clock + *Npas2*			Immediately arrhythmic
	Bmal1 (*Mop3*)	Targeted disruption	Null	Immediately arrhythmic
Period	*Per1*	Targeted disruption	Null	Short period, gradual arrhythmicity
	Per2	Targeted disruption	Deletion	Short period, gradual arrhythmicity
	*Per2**	Missense (S662G)	Dominant-negative	Short period, advanced phase
	Per3	Targeted disruption	Null	Slightly short period
	Per1 + *Per2*			Immediately arrhythmic
	Per1 + *Per2* + *Per3*			Immediately arrhythmic
Cryptochrome	*Cry1*	Targeted disruption	Null	Short period
	Cry2	Targeted disruption	Null	Long period
	Cry1 + *Cry2*			Immediately arrhythmic
Casein Kinase	*CKIδ**	Missense (T44A)	Gain-of-function	Short period, advanced phase
	CKIδ	Targeted disruption	Deletion	Lethal when homozygous; tissues long period.
	*CKIε***	Missense (R178C)	Gain-of-function	Short period, advanced phase
	CKIε	Targeted missense	Gain-of-function	Short period, advanced phase
	CK1ε	Targeted disruption	Null	Slightly longer period/no effect
F box proteins (ubiquitin ligases)	*Fbxl3*	Missense (C358S)	Null	Long period
	Fbxl3	Missense (I364T)	Null	Long period
Orphan Nuclear Receptor	*Rev-erb-alpha*	Targeted disruption	Null	Short period
	RorA	Deletion (Staggerer)	Null	Rhythm instability

Mutations and mutant phenotypes listed were studied in mice, except as noted by asterisks: * indicates Human, ** indicates Syrian hamster. Missense mutations are those that lead to a change in a single amino acid residue; the residue and the change are indicated in single letter format: S662G indicates that a Serine residue at position 662 of human PER2 is mutated to Glycine.
Alleles designated as "Deletion" produce some protein products, but the deletion of genomic material results in deletion-mutant protein products that appear to lack functional activity. From a functional point of view, these are viewed as null alleles. "Null" alleles are those where there is either no product or the product is so truncated that it is irrelevant. "Dominant-negative" alleles result in protein products that are altered in activity, such that the mutant product interferes with the activity of the product from the normal allele, thus producing a mutant phenotype in heterozygotes and a more severe phenotype in homozygotes (semidominant inheritance). "Gain-of-function" alleles are hyperactive in performing the function of the mutated gene. Additional genes not listed may modify the circadian feeback loop in more subtle ways. Results of overexpression studies are not included.

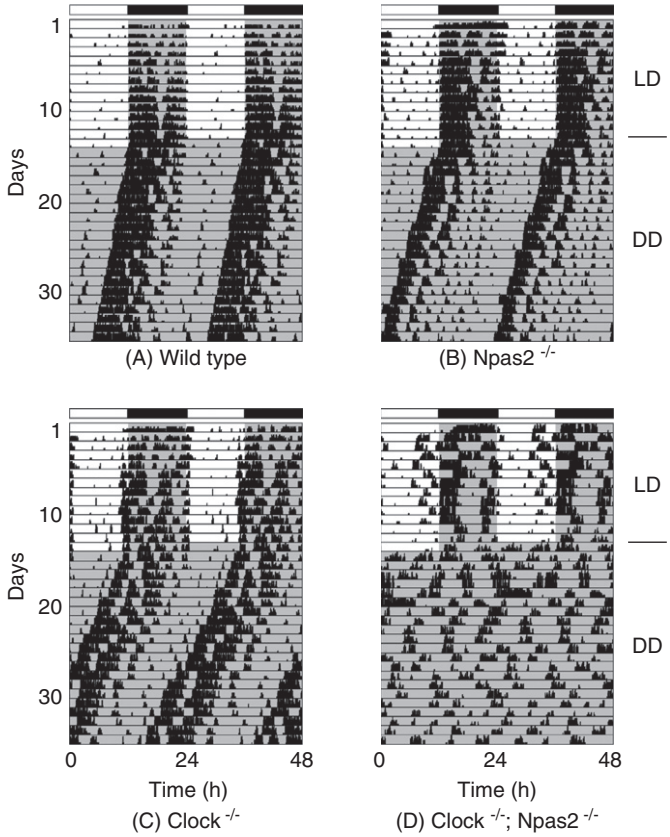


FIGURE 39.8 Molecular redundancy in the maintenance of mouse circadian rhythms. Panels A–D show double-plotted actograms from (A) a wild-type mouse, (B) an NPAS2-deficient mouse, (C) a CLOCK-deficient mouse, and (D) a double-mutant (CLOCK-deficient, NPAS2-deficient) mouse. The animals were initially housed in a light:dark cycle (LD), and then the lights were disabled and they entered constant darkness (DD). Note that the double-mutant mouse becomes arrhythmic in DD, while the mice of the other genotypes "free run" with a cycle length of less than 24 hours. The apparent rhythmicity when the double-mutant mouse is housed in LD is due to partial suppression of locomotor activity by light, a phenomenon called negative masking. Plotting conventions as in Fig. 1. *Modified from DeBruyne, Weaver, & Reppert, 2007, with permission from Nature Neuroscience.*

while null mutations of CK1ε have only a very modest effect. Pharmacological inhibition of both CK1δ and CK1ε greatly lengthens period, indicating compensation by one kinase in the absence of the other is occurring in the genetic models. "Activating mutations" of CK1δ or CK1ε shorten cycle length. Mutation or inhibition of ubiquitin ligases (Fbxl3 or β–TrCP) lengthens period.

The preceding discussion addresses the phenotype of intact mutant mice, using behavioral and physiological rhythms to assess the functional state of the SCN circadian clock. Notably, however, studies of dissociated SCN neurons reveal that a more severe phenotype is observed at the level of single SCN neurons from knockout mice, relative to SCN explants maintained under similar conditions or the *in vivo* phenotype (Liu et al., 2007; Welsh et al., 2010). A small population of weakly functional oscillators apparently can use the network properties of the SCN to generate coherent rhythmicity (DeBruyne, 2008; Welsh et al., 2010).

Clock-Controlled Genes: From the Core Clockwork to Cellular Rhythms

Genes that participate in the core negative feedback loop and whose gene family are necessary for its function are called "clock genes." The impact of the molecular feedback loop occurs through the rhythmic regulation of "clock-controlled genes," which are distinguished from clock genes by the fact that they do not feed back on the core oscillatory mechanism in a manner that makes them necessary for rhythmicity. Some clock-controlled genes with E-box elements in their regulatory regions are controlled in the same manner as *Pers* and *Crys*: their transcription is directly activated by CLOCK:BMAL1 and inhibited by the activity of the PER/CRY complexes disrupting this transcriptional activation. The neuropeptide arginine vasopressin (AVP) is a prototype for this mechanism (Jin et al., 1999). The rhythmic expression of AVP mRNA in the SCN leads to a high-amplitude circadian rhythm of AVP peptide in the cerebrospinal fluid that is insulated from the more well-known osmotic regulation of AVP in blood (see Chapters 37 and 38). The rhythmic synthesis and secretion of AVP from SCN neurons may modulate electrical activity of nearby neurons, increasing the amplitude of the ensemble SCN neuronal firing rhythm and thus affecting many rhythms. AVP of SCN origin also appears to play a role in regulation of the hypothalamo-pituitary-adrenal axis (Chapter 38).

Notably, products of some genes regulated by the CLOCK:BMAL1 complex are themselves able to regulate transcription through additional regulatory elements. For example, a family of D-element binding proteins (DBP/HLF/TEF and E4BP4) are controlled via E-box elements, and these proteins compete for occupation of elements in responsive genes. Regulation of *Rev-erb-α* and *Rora*, noted previously, also occurs through E-box mediated activation. Thus, the rhythmic regulation of these transcription factors, and the multiple regulatory element combinations that can occur in responsive genes, results in a cascade leading to rhythmic gene expression with many possible peak times of expression and with tissue-specific regulation (Ueda et al., 2005).

Summary

The basis for circadian oscillation is a molecular feedback loop. A phase of transcriptional activation alternates with a phase of transcriptional repression, and the cycle length of this rhythm is approximately 24 hours. The bHLH-PAS proteins CLOCK and NPAS2 heterodimerize

with BMAL1 to produce E-box-dependent transcriptional activation. Among the genes controlled primarily by E-box regulation are *Per1* and *Per2*; their high-amplitude temporal regulation is critical to determining the timing of transcriptional repression. PER protein abundance regulates the formation of inhibitory complexes containing PERs, CRYs, and casein kinase proteins. These complexes recruit transcriptional repressors and negate the activity of the activator complex. Post-translational events including phosphorylation and ubiquitination of PERs and CRYs are important for building a time delay into the feedback loop; altering these activities alters circadian cycle length. In parallel with the regulation of "clock genes," the molecular feedback loop regulates the expression of effector genes and also of transcription factors. Through these concentric waves of transcriptional activity, the circadian clock can affect the expression of key (often rate-limiting) genes, in a tissue-specific manner.

CIRCADIAN PHOTORECEPTION IN MAMMALS

In the following sections, we discuss the mechanisms by which the circadian clock in the mammalian SCN is synchronized to the daily light:dark cycle.

Entrainment of the SCN Pacemaker Occurs through Daily Phase Shifts

The intrinsic period of the circadian pacemaker is revealed by measuring it in constant environmental conditions, or in animals made insensitive to light by disruption of anatomical pathways from the eye to the SCN. Rhythms that are not entrained (e.g., not synchronized by a rhythmic environmental cue) are referred to as "free-running" rhythms. The period length of rhythms that are free-running is usually not exactly 24 hours (Fig. 39.1). Instead, the circadian oscillator has a species-typical intrinsic period. Rats housed in constant darkness have a free-running period of approximately 24.3 hr, while mice under the same conditions free-run with a period of approximately 23.6 hr. There are strain variations within these species as well.

For an animal to entrain to a light:dark cycle that has a 24-hour cycle length, light must shift the animal's clock each day by an amount equal to the difference between their intrinsic, free-running period and the 24-hour period of the lighting cycle. While it might seem simpler for the clock to have a cycle length of exactly 24 hours (and forget about daily readjustments), there is seasonal variation in lighting conditions in natural environments, most notably a seasonal change in the length of the light period each day. A timekeeping system that tracks dusk and dawn each day is needed for accurate entrainment under these circumstances.

In assessing the response of the circadian system to light, it is often more informative to assess the response of animals to brief light pulses rather than to full light-dark cycles. A brief pulse of light is provided when an animal is free-running in constant conditions, allowing control over the timing, intensity, and even wavelength of the stimulus. Experiments of this type reveal that the response of the circadian system to light depends strongly on the time of the day at which the light exposure occurs (Fig. 39.9). In constant conditions, the time when the animal "thinks" it is night (as reported by its locomotor activity rhythm) is referred to as "subjective night," while the time that corresponds to internal daytime is "subjective day." Light falling early in the subjective night leads to a phase delay—for example, activity on subsequent days begins at a later time, consistent with a light-induced shift in the timing of the oscillator (Fig. 39.9C). Conversely, light exposure late in the night will cause a phase advance (Fig. 39.9D, E). During large portions of the animal's subjective day, light exposure does not lead to a phase shift (Fig. 39.9A). Plotting the response to brief pulses of light against the time of the light exposure leads to a phase response curve (PRC; Fig. 39.9F). In general, the shape of the PRC to brief light pulses is very similar across species, whether nocturnal, diurnal, or crepuscular, and regardless of whether their intrinsic period is less than or greater than 24 hours. With very strong stimulation (e.g., long duration or bright light exposure) the clock can be reset to a specific point in the circadian cycle (so-called "type 0 resetting"). Many pharmacological stimuli also cause phase-dependent phase shifts of the SCN oscillator, and the shape of the PRC to nonphotic stimuli differs from the PRC to light.

A Specific Retinal Photoreceptive System Mediates Circadian Entrainment to Light

In mammals, detection of light by the circadian system requires the presence of the retina. A direct anatomical connection from the retina to the SCN was first identified in the early 1970s. This retino-hypothalamic tract is necessary and sufficient for photic entrainment of circadian rhythms.

Remarkably, the retinal photoreceptors required for visual image formation are not necessary for circadian photoreception. Studies of mutant mice show that circadian sensitivity to light persists even with virtually complete loss of the rods and cones. Other nonvisual responses to light (inhibition of locomotor activity, suppression of nocturnal melatonin secretion, and constriction of the pupil in response to light) also occur in the absence of rods and cones. These findings suggested that a nonrod, noncone photoreceptive cell mediates these "nonvisual" responses to light.

An opsin-like pigment, melanopsin, was initially identified as a putative photopigment in the light-responsive

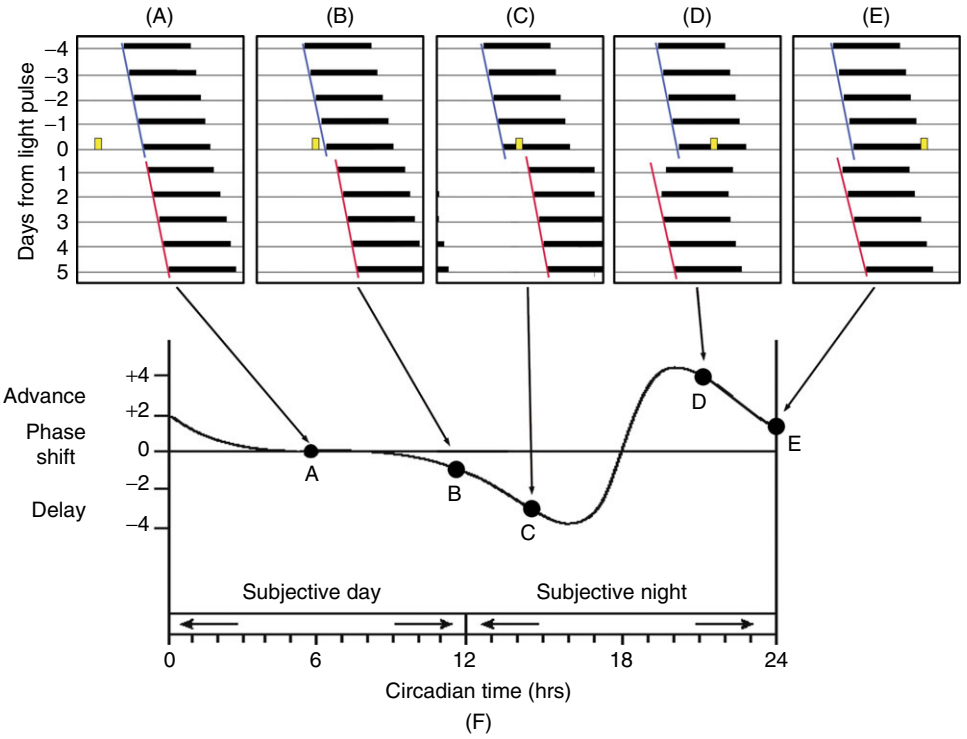


FIGURE 39.9 Phase-dependent effects of light on circadian rhythmicity. (A–E) Schematic representation of the phase-dependent phase shifts that occur following exposure of free-running animals to brief pulses of light. Each panel represents an actogram of one animal housed in constant darkness and exposed to a single 1-hour light pulse (yellow box). The phase shift is the difference, on the day of the light pulse, between a line drawn through activity onsets before the light pulse (blue lines) with a line through activity onsets after the light pulse (red line). In some cases, the behavioral shift is not complete on the first cycle after the light exposure, and the shift seems to occur over several days. The intermediate activity onsets are called "transients" and are ignored when determining the phase after the shift. (F) Phase Response Curve. The phase shift induced by a brief light pulse is highly dependent on the time at which the light exposure occurred. Circadian Time (CT)12 is defined as the time of activity onset in a nocturnal animal. Light exposure occurring during the biological daytime (CT0–12) does not cause a shift. Responses occur only during the animals' biological nighttime. Early in the subjective night, light exposure delays the clock, so activity begins at a later time on subsequent cycles. Late in the subjective night, light exposure causes a phase advance, as detected by earlier activity onsets on the days after the light pulse. Phase-dependent phase shifts allow the circadian oscillator to be entrained to environmental cycles.

melanophores of frog skin, and was later found to be expressed in a subset of retinal ganglion cells (RGCs) in the mammalian retina (Provencio, Jiang, De Grip, Hayes, & Rollag, 1998; Provencio, 2011). More recently, melanopsin has been shown to contribute to circadian photoreception in the mammalian retina. Melanopsin is expressed in a relatively small population (~2400) of RGCs in each mouse eye (Fig. 39.10A). Melanopsin-containing retinal ganglion cells respond directly to light, even when isolated from the rods and cones. Their intrinsic photosensitivity requires the presence of melanopsin.

The population of intrinsically photosentitive RGCs (ipRGCs) is heterogeneous in laminar distribution of dendritic arborization, cellular morphology, level of expression of melanopsin, central projections, and also in function (Schmidt, Chen, & Hattar, 2011). Five subtypes (M1–M5) are recognized (Ecker et al., 2010). Melanopsin can be detected in the M1 subtype by immunohistochemistry (Fig. 39.10B). Mice in which lacZ is inserted into the melanopsin locus provide another way to view melanopsin-expressing cells, although beta-galactosidase staining is detectable only in the M1 subtype. This blue stain is also readily detected in the M1 cells' projections into the central nervous system (Fig. 3.10C,D). The M1's densely innervate the SCN, the intergeniculate leaflet (IGL), and the olivary pretectal nucleus (OPN), a site critical for pupillary constriction in response to light. Additional projections of the M1 class provide innervation to the ventrolateral preoptic area, important in the regulation of sleep, and to the ventral subparaventricular zone (vSPZ), implicated in the inhibition of locomotor activity by light. The M1 class (~700 RGC's) can be further subdivided based on expression of the transcription factor Brn3b. The Brn3b-negative M1 group (~200 cells) is critical for circadian photoentrainment (Chen, Badea, & Hattar, 2011). The Brn3b-expressing M1's mediate pupillary constriction in response to light (Güler et al., 2008). Remarkably, ipRGCs are sufficient for visual pattern discrimination and photic induction of immediate-early gene expression in visual cortex, even in the absence of rod-cone signaling (Ecker et al., 2010).

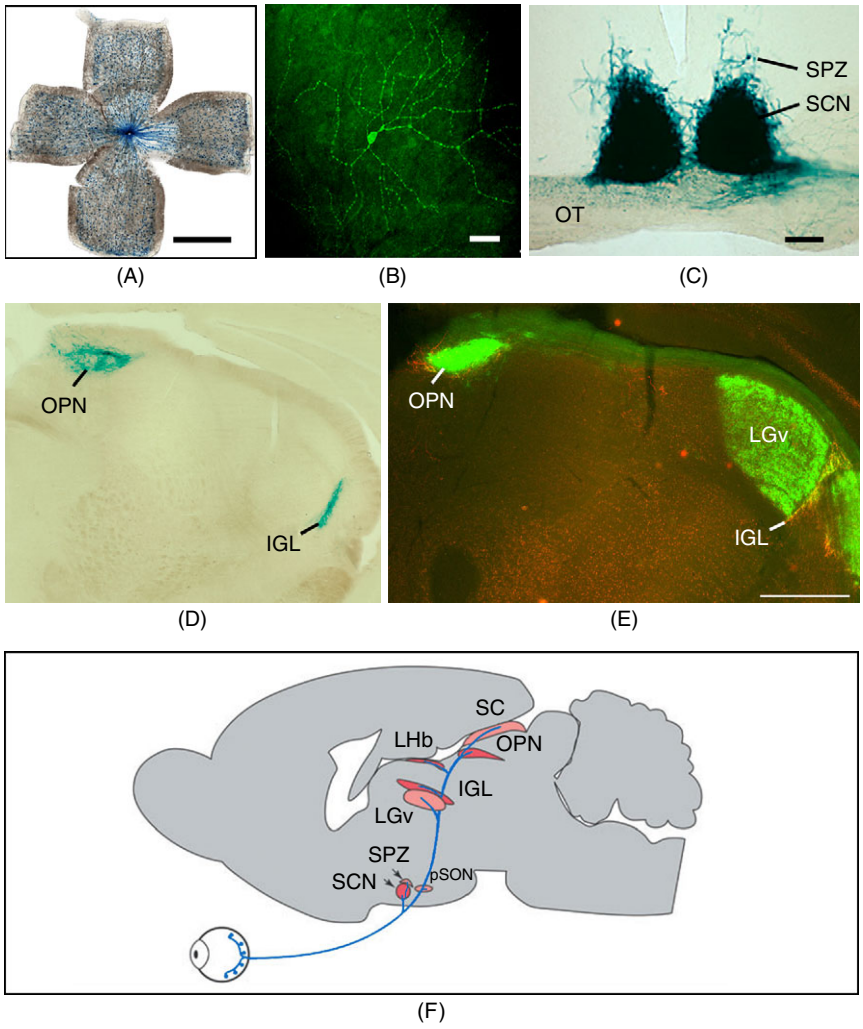


FIGURE 39.10 Retinal ganglion cells containing melanopsin innervate the SCN and other "nonvisual" targets. (A) Melanopsin-expressing retinal ganglion cells. Melanopsin-expressing retinal ganglion cells were detected in this whole-mount preparation of a postnatal day 5 mouse retina using a beta-galactosidase reporter gene inserted into the melanopsin locus. Tissue was processed for histochemical detection of beta-galactosidase activity, resulting in deposition of blue reaction product in cells and processes containing melanopsin. Note the sparse distribution of labeled retinal ganglion cells (individual blue dots); axons course toward the head of the optic nerve in the center of the field. Scale bar is 1 mm. *Images in panels A, C, D, E, and F provided by Samer Hattar, Johns Hopkins University.* (B) Immunofluorescence image of a ganglion cell in the human retina stained for melanopsin. Melanopsin immunoreactivity appears green. Note the large dendritic field of the cell, the localization of melanopsin throughout the dendritic tree, and dendritic varicosities. These cells are ill-suited for spatial localization, but are ideally suited for luminance detection because the entire cell is photosensitive. Scale bar = 100 microns. *Image provided by Ignacio Provencio, University of Virginia.* (C) The SCN receives a strong, bilateral input from melanopsin-containing RGC's. Central projections of melanopsin-expressing retinal ganglion cells were detected using a beta-galactosidase reporter gene inserted into the melanopsin locus as described in Panel A. Melanopsin-containing cell bodies are present only in the retina, so the blue staining seen in the brain represents axonal projections of these retinal neurons. The right eye of the animal had been removed over 2 weeks prior to staining. Labeled, retinal fibers in the optic tract (OT) have degenerated on the apparent left side of the image, leaving weak, granular labeling. The staining in the SCN remains strong on both sides, indicating the extensive, bilateral projections to the SCN. Melanopsin-containing processes also extend into the subparaventricular zone (SPZ). Scale bar = 100 microns. (D) Central projections of melanopsin-containing retinal ganglion cells, revealed by beta-galactosidase staining. The olivary pretectal nucleus (OPN) and intergeniculate leaflet (IGL) receive a crossed input from melanopsin-expressing retinal ganglion cells. (E) Melanopsin is localized within a subset of retinal projections. The anatomical level is comparable to Panel D. Projections of the entire population of retinal ganglion cells were revealed by intraocular injection of cholera toxin B subunit three days prior to tissue collection, allowing anterograde labeling of the fibers of all retinal ganglion, followed by detection of the toxin which appears green. Note the intense labeling in the dorsal lateral geniculate nucleus (LGd). Melanopsin expression in retinal terminals was detected by red immunofluorescence; in this merged image, the overlap of red and green channels in the melanopsin-containing retinal ganglion cells yields yellow labeling in the IGL and OPN. Scale bar (for panels D and E) = 500 microns. (F) Schematic illustration of the axonal projections of the M1 class of melanopsin-expressing retinal ganglion cell as revealed by beta-galactosidase staining. Principle targets (SCN, OPN, IGL, and lateral habenula [LHb]) are indicated by dark red shading. Less extensively innervated structures indicated by light red shading are the SPZ, superior colliculus, peri-supraoptic nucleus, and ventral subdivision of the lateral geniculate nucleus. Minor targets (not shown) include the preoptic area, bed nucleus of the stria terminalis, medial amygdaloid nucleus, anterior hypothalamic area, lateral hypothalamus, LGd, and periaqueductal gray. (See Hattar *et al.*, 2006.)

The phototransduction mechanism in melanopsin-containing cells appears to be more similar to the invertebrate visual signaling cascade (in which phosphoinositide signaling activates cation channels of the transient receptor potential channel type), rather than through the typical vertebrate photoreceptor cascade involving transducin, cGMP, and cyclic nucleotide gated ion channels.

Despite being well positioned to serve as the circadian photoreceptive molecule, melanopsin is in fact not the only photopigment involved in circadian photoreception. Melanopsin-deficient mice have a reduction in the effects of light on phase shifting, negative masking of locomotor activity, and pupillary responses at high light intensities, but these responses to light are not completely absent (Provencio, 2011). Importantly, mice that lack rod and cone function and also lack melanopsin appear completely blind with respect to both image formation and nonimage forming responses to light (Fig. 39.11). Rods and cones can signal through the melanopsin-expressing RGC's, even if the latter cells are melanopsin-deficient. Destruction of the melanopsin-expressing cells, however, blocks nonvisual photoreception (Güler et al., 2008; Hatori et al., 2008). Thus, rod-cone input converges on the melanopsin-containing retinal ganglion cells, and parallel pathways that control different nonvisual image formation functions arise from defined subsets of these ipRGCs (Schmidt et al., 2011).

Retinal Input to the SCN: Anatomy and Neurochemistry

The melanopsin-expressing retinal ganglion cells that project to the SCN use glutamate as their primary neurotransmitter. A large percentage of these retinal neurons also contain the neuropeptide Pituitary Adenylate Cyclase Activating Peptide (PACAP). Application of these agents to SCN slices *in vitro* causes phase-dependent phase shifts of intrinsic SCN rhythms. Furthermore, glutamate receptor antagonists block light-induced phase shifts *in vivo*, revealing the functional importance of glutamatergic input to the SCN.

Axon collaterals of the melanopsin-positive retinal ganglion cells innervate neurons in the intergeniculate leaflet (IGL) that express neuropeptide Y (NPY). These NPY-containing neurons project to the SCN as the geniculohypothalamic tract. The geniculohypothalamic tract is thought to play a major role in nonphotic phase shifting, a phenomenon in which behavioral arousal (usually accompanied by intense locomotor activity) during the circadian day leads to increased NPY release within the SCN and leads to a phase shift of activity rhythms. The indirect retinal pathway through the intergeniculate leaflet also appears to modulate photic entrainment. The direct retino-hypothalamic tract and the indirect retino-geniculo-hypothalamic pathway converge in the SCN, in the "core" region defined by the dense retinal input and by a subpopulation of neurons expressing vasoactive intestinal peptide (VIP). Another major input to the SCN core is a serotonergic input from the raphe nuclei, which modulates the impact of light input to the SCN through presynaptic serotonin receptors on retinohypothalamic tract terminals.

Molecular Mechanisms of Entrainment: Light-Induced *Per* Expression

Per1 and *Per2* gene expression levels are rhythmic in the SCN, with high levels during the daytime (Fig. 39.7). These molecular rhythms persist in constant darkness. Exposure

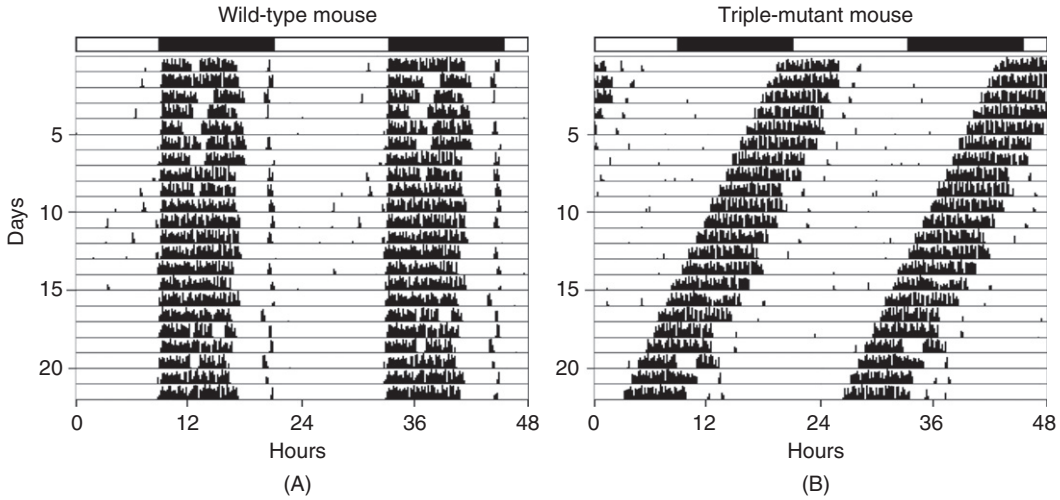


FIGURE 39.11 Rods, cones, and intrinsically photosensitive retinal ganglion cells expressing melanopsin mediate circadian entrainment. (A) is a double-plotted actogram of a control mouse showing normal entrainment to the light:dark cycle. (B) shows a double-plotted actogram from a triple-mutant mouse, lacking rod and cone phototransduction mechanisms and also lacking melanopsin. Triple-mutant animals free-run as though there is no lighting cycle. Triple-mutant animals also lack pupillary constriction responses and acute inhibition of locomotor activity by light (Hattar et al., 2003). The disruption of rod and cone signaling is accomplished by breeding to generate animals homozygous for disruption of the alpha subunit of rod transducin and the cone cyclic nucleotide gated channel A3 subunit. *Actograms provided by Samer Hattar, Johns Hopkins University.*

to light at night causes a large, transient increase in both *Per1* and *Per2* transcript levels. *Per* gene expression in the SCN is unresponsive to light exposure during the subjective daytime, when light exposure does not cause phase shifts. The temporally "gated" induction of *Per1* and *Per2* gene expression is likely mediated by activation of MAP kinase and CREB phosphorylation by retinal input, leading to induction of *Per1* and *Per2* expression through Ca^{2+}/cAMP response elements (CRE) in the promoters of these genes. (A third *Period* gene, *Per3*, lacks a promoter CRE element and is unresponsive to light even at night. *Per3* has, at most, a very modest contribution to the core circadian oscillatory mechanism or the resetting mechanism).

Increased *Per1* and *Per2* expression are the only widely reproducible changes in clock gene expression that occur in response to light, focusing attention on these genes as important transducers of the effects of light on the molecular clock. Studies in mutant mice reveal that neither gene is necessary for phase shifting responses to light, however (Pendergast et al., 2010). *Per1* and *Per2* may be redundant in function for phase shifting, with either gene able to mediate the re-setting response in the absence of the other.

Peak nuclear PER1 and PER2 protein levels occur at the time of the light-to-dark transition (Fig. 39.7). Light-induced increases in PER1 and PER2 protein early in the night would extend the period of transcriptional inhibition by the PER/CRY complex. This may be the mechanism by which light exposure early in the night causes a phase delay. In contrast, light-induced PER expression late in the night would lead to an earlier-than-normal initiation of PER/CRY-mediated inhibition, resulting in an advance of SCN molecular rhythmicity in subsequent cycles. Thus, light-induced alterations of PER expression and interaction of the light-induced PER with the molecular feedback loop may explain the phase-dependent phase shifting effect of light shown in Figure 39.9.

While light does not influence *Per* levels during the animal's subjective daytime and does not cause phase shifts at this time, nonphotic manipulations can cause phase shifts when administered during the daytime (e.g., novelty-induced voluntary locomotor activity, arousal, serotonin receptor activation). These manipulations lead to *reduced Per* gene expression in the SCN. Thus, acute regulation of *Per* gene expression may be the mechanism by which a variety of manipulations affect circadian phase.

Summary

The circadian clock in the SCN is entrained to the daily light-dark cycle by light causing phase-dependent phase shifts each day. While the retinae are required for entrainment, the retinal photoreceptors necessary for visual image formation, the rods and cones, are not needed. Instead, a population of intrinsically photoreceptive retinal ganglion cells conveys information about the level of light to the SCN and to other brain regions that regulate nonvisual photoreception. The photopigment in these retinal ganglion cells is melanopsin, and they utilize glutamate and PACAP as their transmitters. In the absence of melanopsin, rods and cones appear capable of signaling through the ipRGC's to result in nonvisual responses to light, including entrainment. Specific subsets of melanopsin-expressing cells have distinct functions. The impact of nocturnal light exposure on SCN physiology includes numerous signaling responses and induction of *Per* gene expression. Light-induced alterations in *Per* gene expression likely provide a key role in resetting of the molecular feedback loop within SCN neurons.

CIRCADIAN OUTPUT MECHANISMS

From Molecular Feedback Loops to Firing Rate Rhythms

The molecular feedback loop within individual SCN neurons ultimately has its impact on physiological processes through regulation of SCN neuronal firing rate (Kuhlman & McMahon, 2006). Rhythmic regulation of potassium channels in the membrane of SCN neurons appears to be critical for producing circadian rhythms in membrane potential and firing rate.

Large-conductance, calcium-activated potassium channels (BK channels) are rhythmically expressed in the SCN, with high levels of transcript levels and BK channel activity during the night contributing to the reduction in firing rate at night (Meredith et al., 2006). Disruption of the BK channel (*Kcnma1*$^{-/-}$) results in higher SCN firing rate selectively at night, and the mutant mice have greatly reduced amplitude of several circadian output rhythms.

In addition, the fast delayed rectifier (FDR) current is critical for robust circadian rhythms in behavior. The Kv3.1 and Kv3.2 potassium channels and their activity are more abundant in the SCN during the circadian daytime, and blocking their activity reduces SCN rhythmicity (Itri, Michel, Vansteensel, Meijer, & Colwell, 2005). Mice lacking both the *Kcnc1* and *Kcnc2* genes do not express the Kv3.1 and 3.2 channels in the SCN, have greatly reduced FDR current, and have severely disrupted circadian behavioral rhythms (Kudo, Kuljis, Constance, & Colwell, 2011). It thus appears that multiple ionic mechanisms contribute to rhythmicity in SCN neuronal firing rate, and regulation of channel activity is probably accomplished through both transcriptional and post-transcriptional mechanisms.

Rhythmic firing of SCN neurons not only serves as its mechanism of communication with downstream target regions, but also plays an important role in synchronizing the cellular oscillators within the SCN to each other. When

sodium-dependent action potentials are blocked for several days, SCN neurons become desynchronized from one another. The mechanism for this may be as simple as the loss of rhythmic, synaptic release of neurotransmitters, including VIP. Indeed, synaptic vesicle cycling plays an important role in the regulation of circadian rhythms (Deery et al., 2009).

Humoral Outputs from the SCN Also Influence Locomotor Activity Rhythms

The rhythmic electrical activity of SCN neurons likely leads to rhythmic neurotransmitter release and modulation of downstream, synaptic targets along defined anatomical pathways as discussed in greater detail below. Several lines of evidence indicate that the SCN also influences locomotor activity through secretion of humoral output molecules. Among these lines of evidence are the findings that (1) transplants of fetal SCN tissue that restore locomotor activity rhythms do not restore other rhythms, including endocrine rhythms, and (2) transplanted SCN tissue encapsulated so as to allow diffusion of small molecules while preventing axonal outgrowth nevertheless can restore locomotor activity rhythms in SCN-lesioned hamsters (Silver, LeSauter, Tresco, & Lehman, 1996). The latter finding suggests that substances secreted by the SCN act in regions near the SCN (but not in the SCN itself) to regulate locomotor activity. Three substances proposed to play a role in regulating locomotor activity rhythms in rodents are transforming growth factor alpha, cardiotropin-like cytokine 1, and prokineticin 2.

Summary

For oscillations in the abundance of circadian clock gene products in SCN neurons to affect physiology and behavior, the molecular feedback loop must impact on more overt aspects of cellular rhythmicity. One mechanism for this to occur is through circadian regulation of transcripts encoding ion channels; indeed, potassium channels have been identified that contribute to circadian rhythmicity. Through regulation of potassium channel activity, rhythmic transcription factor activity is translated into rhythmicity in neuronal firing rate and neurotransmitter release. While much of the output of the SCN comes via synaptic interactions to downstream targets, substances released from the SCN by a humoral mechanism can also influence locomotor activity.

DIVERSITY OF OUTPUT PATHWAYS LEADING TO PHYSIOLOGICAL RHYTHMS

Animals that lack circadian rhythms (those with destruction of the SCN or genetic disruption of key circadian clock genes) survive, and these animals generally can reproduce. Nevertheless, the circadian timing system is thought to optimize "internal temporal order" by coordinating rhythms in different tissues, and to allow anticipation of physiological demands imposed by the predictably changing environment. The impact of the circadian oscillator is exerted through the rhythmicity it imposes on physiological functions.

The following sections address the regulation of several physiological rhythms. In most cases, we emphasize how these rhythms are "wired" to the SCN through neuroanatomical pathways and emphasize rhythms for which the SCN can be viewed as a driving pacemaker. It is important to note that SCN-generated physiological rhythms can also entrain peripheral oscillators, without a direct SCN-to-target output (Kornmann et al., 2007; Vollmers et al., 2009; Buhr, Yoo, Takahashi, 2010; see Fig. 39.3). Subsequent tissue-autonomous rhythmicity can profoundly influence physiology and response to perturbations (Box 39.4) (Kornmann et al., 2007; Marcheva et al., 2010; Sadacca, Lamia, deLemos, Blum, & Weitz, 2011)

Amplification of SCN Output and Divergence of Output Pathways

Considering the pervasive influence of the SCN on rhythmic physiology, the direct anatomical output of the SCN is surprisingly limited. SCN efferents project heavily to a column of cells dorsal and caudal of the SCN, including the subparaventricular zone (SPZ) and the dorsomedial nucleus of the hypothalamus (DMH) (Figs. 39.12 and 39.13). Neurons in the SPZ and DMH in turn project to brain regions that ultimately regulate rhythmic activities (Fig. 39.13). Many effector regions receive a small, direct innervation from the SCN that converges with the larger, indirect output via the SPZ. The SPZ thus appears to relay and reinforce the direct output of the SCN, and allows convergence of influences from other brain areas. Lesions that kill cell bodies in the ventral SPZ (while leaving axons passing through the region intact) disrupt the circadian rhythms of sleep/wakefulness and locomotor activity (Lu et al., 2001). Amplification of SCN output in the vSPZ is thus necessary for these rhythms; direct SCN output to target sites is insufficient. In a similar manner, fiber-sparing lesions of the dorsal part of the SPZ (dSPZ) selectively disrupt the body temperature rhythm while leaving sleep and activity rhythms intact.

Another level of amplification and integration occurs in the DMH, which is heavily innervated by the ventral SPZ as well as receiving direct innervation from the SCN (Figs. 39.12 and 39.13). Lesions of the DMH specifically disrupt circadian rhythms of sleep, locomotor activity, corticosterone secretion and feeding, but not body temperature (Chou et al., 2003).

BOX 39.4

CHRONOPHARMACOLOGY

Many drugs, toxins, and stressors have effects that vary over the 24-hour day. Many of these variations represent true circadian rhythms in responses—for example, they persist in constant environmental conditions. Chronopharmacology is the subdiscipline of chronobiology that seeks to identify and understand these rhythms, and it is an important field, since altering the time of drug administration can have a profound impact on its efficacy, toxicity, and the ratio between these two measures (the therapeutic index). For this reason, there is often a specific time for optimal response to a chemotherapeutic agent, and this optimal timing varies between agents depending on their mechanism of action. Rhythmicity in drug effects could be due to alterations in bioavailability (absorption, metabolism, or excretion) or in end organ sensitivity. Rhythmic variation in therapeutic index has been well documented for several chemotherapeutic agents and anesthetics. There are also robust rhythms in responses to toxic agents, including infectious agents such as *E. coli* and endotoxin derived from bacterial cell walls. Recent studies have begun to unravel the molecular mechanisms underlying rhythms in detoxification and drug sensitivity, and focus attention on rhythmicity within target cell populations.

David R. Weaver

Further Readings

Gachon, F., & Firsov, D. (2011). The role of the circadian timing system on drug metabolism and detoxification. *Expert Opin Drug Metab Toxicol, 7*, 147–158.

Innominato, P. F., Lévi, F. A., & Bjarnason, G. A. (2010). Chronotherapy and the molecular clock: Clinical implications in oncology. *Advance Drug Delivery Review, 62*, 979–1001.

Lévi, F., & Schibler, U. (2007). Circadian rhythms: Mechanisms and therapeutic implications. *Annual Review of Pharmacology Toxicology, 47*, 593–628.

Lévi, F., Filipski, E., Iurisci, I., Li, X. M., & Innominato, P. (2007). Cross-talks between circadian timing and cell division cycle determine cancer biology and therapeutics. *Cold Spring Harbor Symposia Quantitative Biology, 72*, 465–475.

Paschos, G. K., Baggs, J. E., Hogenesch, J. B., & FitzGerald, G. A. (2011). The role of clock genes in pharmacology. *Annual Review Pharmacology Toxicology, 50*, 187–214.

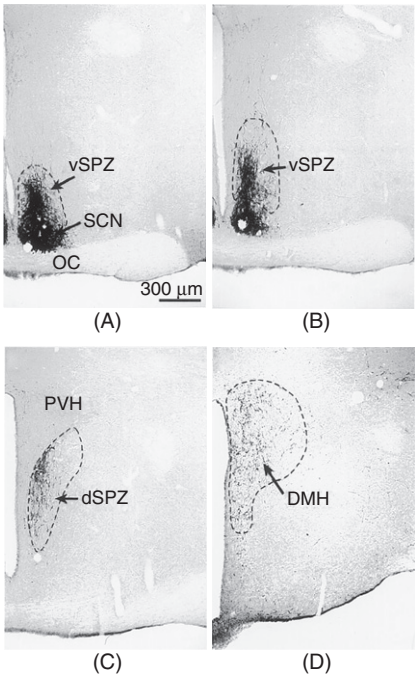


FIGURE 39.12 Distribution of vasoactive intestinal peptide (VIP) immunoreactivity in the rat brain reveals the major output pathways from the SCN. Staining reveals VIP-immunoreactive cell bodies in the SCN and their projections. The major projection of the SCN is to a cell column dorsal and caudal of the SCN, including the subparaventricular zone (SPZ) and the dorsomedial nucleus of the hypothalamus (DMH). (*Reprinted from Lu et al., 2001, with permission from the Society for Neuroscience.*)

Circadian Regulation of Sleep and Wake

Among the most widely recognized roles of the circadian timing system is in the regulation of behavioral state. Our level of alertness is strongly influenced by two biological factors: the time previously spent awake and the circadian time (see Chapter 40). There is a strong and opposing influence of circadian time and the time spent awake, a phenomenon described as the "opponent-process" model for behavioral state control. With increasing time awake, subjects experience a greater homeostatic drive to sleep. The circadian timing system opposes this drive, by providing an arousing signal that reaches peak levels late in the afternoon/early evening in humans.

The influence of the circadian timing system on sleep is mediated by SCN efferents that indirectly reach the brain regions regulating sleep. As noted above, a pathway from the SCN via the vSPZ and DMH is necessary for circadian regulation of sleep. In contrast to the sparse connections of the SPZ to sleep-regulatory areas, the DMH is a primary source of input to the ventrolateral preoptic area (VLPO) and to the neurons of the lateral hypothalamus that secrete the neuropeptide orexin, that are critical for regulation of sleep. The output from the DMH to the VLPO is GABAergic, therefore promoting wakefulness by inhibiting the "sleep switch" neurons. The output from the DMH to the orexin neurons in the lateral hypothalamus contains glutamate and thyrotropin-releasing hormone, and is

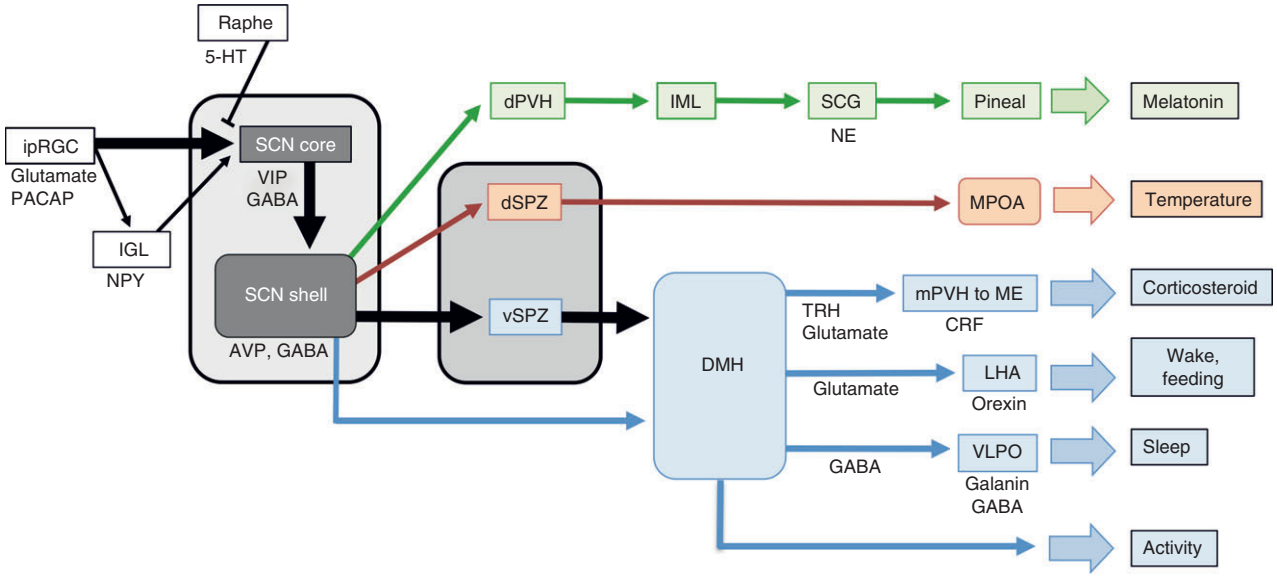


FIGURE 39.13 Schematic of the rodent circadian system, with emphasis on the divergence of output pathways. Rhythmically regulated events are listed on the right, and the pathways leading to their circadian regulation are shown. Major inputs to the SCN are the melanopsin-containing ipRGC's, neuropeptide Y (NPY)-containing neurons in the intergeniculate leaflet (IGL), and serotonergic afferents from the raphe that influence the RHT input. Inputs converge primarily in the core region, although many additional regions project into the SCN (not shown). The SCN core region projects extensively to the shell. The major output of the SCN arises from the shell and is directed caudally to the subparaventricular zone (SPZ). The dorsal subparaventricular zone (dSPZ) projects to the medial preoptic area (MPOA), and is important for circadian regulation of core body temperature. Axons from the ventral SPZ (vSPZ) and direct SCN projections converge in the dorsomedial hypothalamic nucleus (DMH). From the DMH arise distinct projections critical for the circadian regulation of the hypothalamic-pituitary-adrenal axis, sleep/wake and feeding; the next anatomical site along each of these pathways is indicated. The DMH is also critical for regulation of locomotor activity rhythms, but effector structures are not clear. Regulation of the adrenal rhythm and of pineal melatonin secretion occur through the medial parvicellular paraventricular nucleus (mPVH) and dorsal parvicellular PVH (dPVH), respectively. Where known, neurotransmitters are indicated. Abbreviations: LHA, lateral hypothalamic area; VLPO, ventrolateral preoptic area; ME, median eminence; NE, norepinephrine; PACAP, Pituitary adenylate cyclase activating peptide; TRH, thyrotropin releasing hormone.

presumably excitatory, thus promoting wakefulness. While the DMH has few direct projections to the brainstem areas that comprise the ascending arousal system, the orexin neurons have extensive, excitatory projections to these regions (Saper, Lu, Chou, & Gooley, 2005). Thus, a multistage pathway from the SCN to sleep regulatory centers allows integration of other inputs (emotional and cognitive inputs as well as the homeostatic drive for sleep) and promotes sharp transitions between states, rather than direct control of behavioral state. The importance of orexin neurons in coordinating the activity of this system is revealed by narcolepsy, a sleep disorder caused by autoimmune destruction of the orexin-containing neurons in the lateral hypothalamus, and mimicked in animal models by disruption of orexin or orexin receptors (see Chapter 40).

One relatively small subclass of sleep disorders are the circadian-based sleep disorders (Box 39.3). While most insomnias are due to difficulty in initiation or maintenance of sleep, or are related to airway patency during sleep (sleep apnea), circadian-based sleep disorders involve problems with the relative timing of sleep. Transient disruption of sleep often occurs with attempts to shift the relative timing of sleep, as occur in shift work and jet lag (Box 39.5)

Circadian Regulation of Neuroendocrine Functions

The SCN output pathway leading to rhythms in adrenal corticosteroid secretion starts with a projection to the paraventricular nucleus of the hypothalamus (PVN). A subset of neurons in the PVN contain corticotrophin releasing hormone (CRH), and project to the external zone of the median eminence. CRH, released into the pituitary portal system, regulates the release of adrenocorticotropic hormone (ACTH) from the anterior pituitary. ACTH, in turn, is carried in the bloodstream to the adrenal gland, where it stimulates secretion of adrenal steroids. Adrenal steroids feed back to inhibit CRH and ACTH release in a typical neuroendocrine feedback loop (see Chapter 38). This feedback loop is modulated by the circadian timing system, however: the set point for steroid feedback apparently varies over the course of the day (e.g., higher corticosteroid levels are required to reduce ACTH release at some times of day than at others). As a result, corticosteroid levels vary over the course of the day with a sinusoidal rhythm. Light exposure at night can activate the sympathetic innervation of the adrenal gland via the SCN,

BOX 39.5

JET LAG AND SHIFT WORK

Jet lag is induced by flying across several time zones and represents the lack of coordination of the flyer's body to the new local time. Individuals usually entrain to a phase-shifted environment at the rate of about 1 hour per day, although careful control over when light exposure occurs may improve this rate.

Another environmentally imposed disruption of circadian rhythms occurs with shift work. Most people working the night shift remain entrained to the solar day, so that their biological rhythms (e.g., alertness, temperature and hormonal rhythms) are in opposition to the demands of the work schedule. Some of the negative health effects of shift work may be due to the stress caused by trying to sleep when the biological clock dictates it is time to be awake, and trying to be alert and productive when the body expects rest. Clear performance deficits occur when individuals attempt to work for extended periods, ignoring the need for sleep. Performance deficits also occur at specific times of day, irrespective of prior sleep history.

In rodent models, repeated shifting of the light-dark cycle, as occurs with jet lag or shift work, can have significant adverse consequences on health. Circadian clocks normally coordinate internal rhythms to each other and to the environment. Disrupting this coordination may similarly have adverse consequences on human health.

David R. Weaver

Further Readings

Castanon-Cervantes, O., Wu, M., Ehlen, J. C., Paul, K., Gamble, K. L., Johnson, R. L., … Davidson, A. J. (2010). Dysregulation of inflammatory responses by chronic circadian disruption. *Journal of Immunology, 185,* 5796–5805.

Davidson, A. J., Sellix, M. T., Daniel, J., Yamazaki, S., Menaker, M., & Block, G. D. (2006). Chronic jet-lag increases mortality in aged mice. *Current Biology, 16,* R914–916.

Preuss, F., Tang, Y., Laposky, A. D., Arble, D., Keshavarzian, A., & Turek, F. W. (2008). Adverse effects of chronic circadian desynchronization in animals in a "challenging" environment. *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology, 295,* R2034–2040.

Scheer, F. A., Hilton, M. F., Mantzoros, C. S., & Shea, S. A. (2009). Adverse metabolic and cardiovascular consequences of circadian misalignment. *Proceedings of the National Academy of Sciences of the United States of America, 106,* 4453–4458.

Wang, X-S., Armstrong, M. E. G., Cairns, B. J., Key, T. J., & Travis, R. C. (2011). Shift work and chronic disease: the epidemiological evidence. *Occupational Medicine* (Lond), *61,* 78–89.

Martino, T. A., Oudit, G. Y., Herzenberg, A. M., Tata, N., Koletar, M. M., Kabir, G. M., … Sole, M. J. (2008). Circadian rhythm disorganization produces profound cardiovascular and renal disease in hamsters. *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology, 294,* R1675–1683.

bypassing the pituitary and directly increasing corticosteroid production (Ishida et al., 2005).

The SCN also regulates circadian rhythms in other neuroendocrine activities. In rodents, the timing of ovulation is strictly regulated and occurs early in the active period on the day of behavioral estrus (de la Iglesia & Schwartz, 2006). Ovulation is blocked by injection of the anesthetic pentobarbital early in the afternoon on the day of estrus. Remarkably, ovulation then occurs the following day—exactly 24 hours after it should have occurred. Ovariectomized, estradiol-treated female rats also have daily surges of luteinizing hormone in the late afternoon. This is the neuroendocrine equivalent of ovulation, occurring in the absence of the ovaries. These luteinizing hormone surges persist when animals are housed in constant darkness and are not present in animals with SCN lesions. An SCN-dependent neuronal output leads to daily activation of gonadotropin releasing hormone-containing neurons (de la Iglesia, Meyer, Schwartz, 2003).

Twice-daily prolactin surges can be induced in female rats by mating (or by vagino-cervical stimulation that mimics mating, induces pseudopregnancy). The prolactin surges are timed by the SCN. These prolactin surges are ultimately regulated by dopamine released from the median eminence. Feedback of prolactin on the hypothalamic dopamine neurons is also important. VIP neurons in the SCN and oxytocin neurons in the PVN appear to be involved in regulating the hypothalamic dopamine neurons and thus the prolactin surges (Egli, Bertram, Sellix, Freeman, 2004).

Circadian Regulation of Autonomic Function

Many physiological rhythms are regulated by autonomic function. Pathways resulting in rhythmic autonomic activity originate in the SCN, utilize GABA as a neurotransmitter, and generally project to the paraventricular nucleus of the hypothalamus (PVN). Neurons in the PVN (many containing oxytocin) have direct projections to the spinal cord, where they regulate the activity of autonomic preganglionic neurons. Additional pathways include modulation of vagal activity. These pathways regulate

autonomic responses leading to circadian variation in blood pressure and heart rate. These pathways also influence metabolism through direct innervation of fat pads and the liver, the latter of which contributes to the circadian rhythm in plasma glucose levels (Cailotto et al., 2005; Kalsbeek et al., 2006).

Circadian Regulation of Pineal Function

One very well-characterized output rhythm regulated via the autonomic nervous system is the rhythmic production of melatonin. This rhythm is regulated by an anatomical pathway pathway from the SCN via the PVN to the pineal gland (Fig. 39.14A). GABA released from SCN neurons inhibits neurons in the PVN during the day, while the reduction in SCN GABA release at night apparently allows the PVN neurons to be more active, driving melatonin synthesis (Kalsbeek et al., 2000). The PVN neurons project to autonomic preganglionic neurons in the intermediolateral cell column of the spinal cord. The preganglionic neurons project to the superior cervical ganglion, whose fibers innervate the pineal gland. Norepinephrine from these fibers increases the activity of the penultimate enzyme in melatonin biosynthesis, arylalkylamine N-acetyltransferase (AANAT), leading to melatonin synthesis (Fig. 39.14B). AANAT activity is controlled by a combination of transcriptional mechanisms and switch-like regulation of proteasomal degradation of AANAT protein.

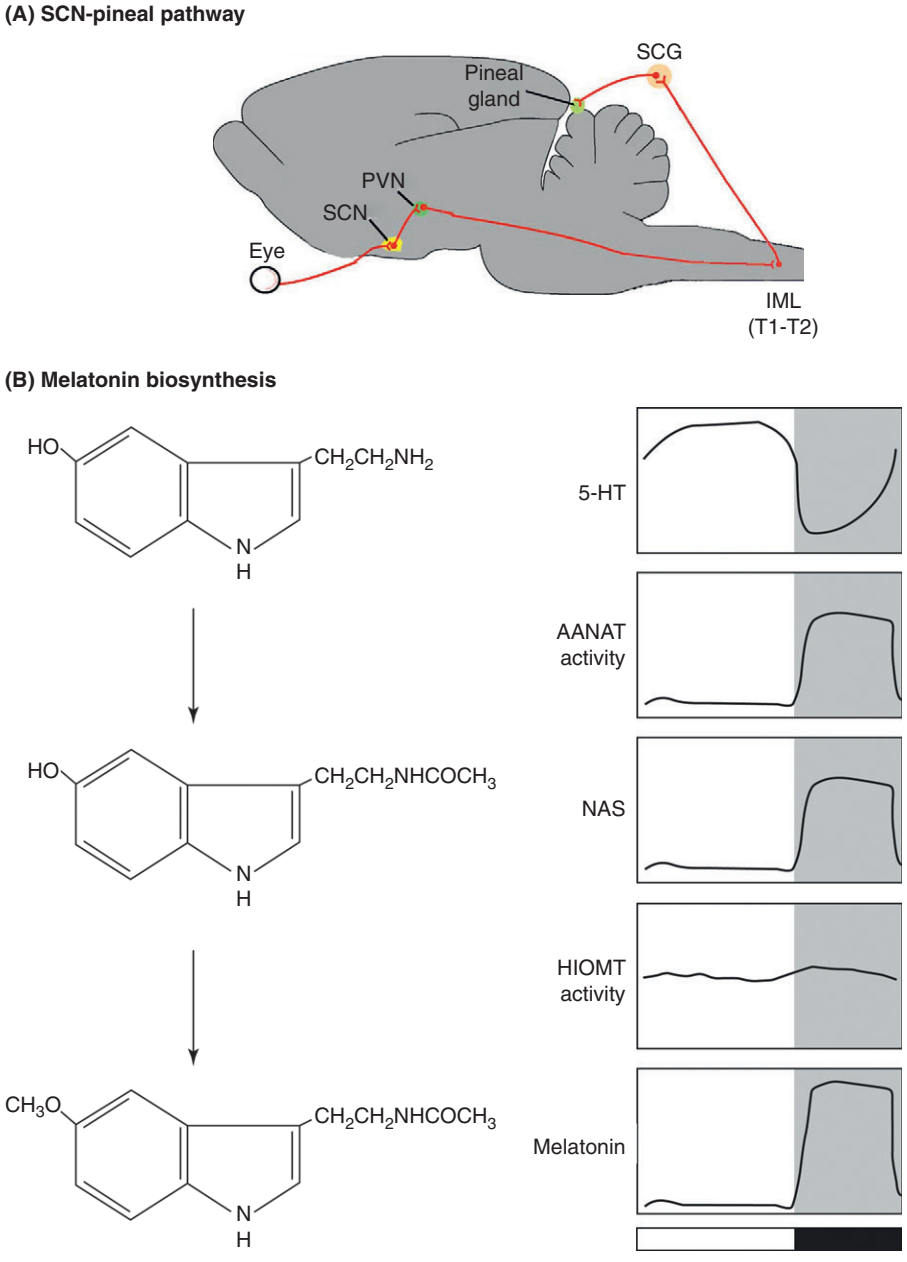


FIGURE 39.14 (A) Schematic illustration of the neuroanatomical circuit regulating pineal melatonin secretion. Photic input detected in the retina is relayed to the suprachiasmatic nucleus (SCN), and then to the paraventricular nucleus of the hypothalamus (PVN). A subset of PVN neurons projects to the intermediolateral cell column (IML) of the spinal cord. Preganglionic sympathetic cell bodies at thoracic levels T1 and T2 project to the superior cervical ganglion (SCG), which innervates the pineal gland. Norepinephrine released from the terminals of SCG neurons is the primary input to the pineal responsible for stimulating melatonin production at night. Light entrains the SCN clock controlling melatonin, but light at night also acutely inhibits melatonin synthesis by disrupting the sympathetic input to the gland. Blood melatonin levels drop within minutes after exposure to light at night. (B) Biosynthesis of melatonin from serotonin (5-hydroxytryptamine, 5-HT). The rate-limiting and highly rhythmic step is the regulation of arylalkylamine N-acetyltransferase activity, which modestly depletes the pineal gland of 5-HT at night as it converts 5HT to N-acetyl-serotonin (NAS). Hydroxy-indole-O-methyltransferase (HIOMT) converts NAS to melatonin (N-acetyl,5-methoxytryptamine). There is no storage mechanism for melatonin; melatonin diffuses into the cerebrospinal fluid and bloodstream as it is produced, making blood melatonin levels a good reflection of the sympathetic input to the pineal gland.

Circadian Regulation of Melatonin Is Critical for Seasonal Reproduction

Several strategies exist to deal with the varying environmental conditions and varying levels of resources that occur over the course of the year. These strategies include seasonal migration to allow highly mobile species to escape challenging conditions (Box 39.6), and, for those species that remain, seasonal regulation of physiology and behavioral adaptation (see Box 39.1).

Many mammalian species restrict breeding activity to specific times of year, resulting in birth of young in the spring and summer months when food is abundant and conditions are most favorable for their survival. Responses to seasonal changes in day length (so-called "photoperiodic responses") include seasonal reproduction, but also extend to an array of physiological responses including seasonal molting of the coat (mediated by seasonal alterations in prolactin secretion), thermoregulatory adjustments (daily torpor and hibernation), alterations in metabolism and body weight, and alterations in behavior. These adjustments improve the probability of survival through challenging times. Like circadian rhythms, these seasonal responses are driven by a system that anticipates predictable environmental changes and prepares for them in a predictive, rather than reactive, manner.

BOX 39.6

LOST WITHOUT A CLOCK

A most fascinating function of circadian clocks is their use in time-compensated sun compass orientation, a phenomenon that was first described by Karl von Frisch in foraging honeybees and by Gustav Kramer in migratory birds. Time-compensated sun compass orientation is also used by eastern North American monarch butterflies (*Danaus plexippus*) during their spectacular fall migration. During the fall journey, the butterflies use their navigation skills to fly south from the United States and southern Canada to overwinter in a few restricted sites atop the Transvolcanic Mountains in central Mexico, some having traveled distances approaching 4000 kilometers.

The amazing navigation abilities of monarch butterflies are part of a genetic program initiated in the migrants, as those that make the trip south are at least two generations removed from the previous generation of migrants. Time-compensated sun compass orientation is believed to be an essential component of navigation required for successful migration. Time compensation is provided by the butterfly's circadian clock. The clock allows the butterflies to continually correct their flight direction, relative to changing skylight parameters, to maintain a fixed flight bearing as the sun moves across the sky during the day (Fig. B39.6A). The importance of the circadian clock in regulating the time-compensated component of flight orientation in monarch butterflies has been shown in two ways. First, clock-shift experiments, in which the light-dark cycle is either advanced or delayed, cause predictable alterations in the direction the butterflies fly (Fig. B39.6B). Second, removing the time compensation clocks abolishes the time-compensated component of flight orientation (Fig. B39.6C). It is thus clear that time compensation is a requirement for successful migration.

Steven M. Reppert

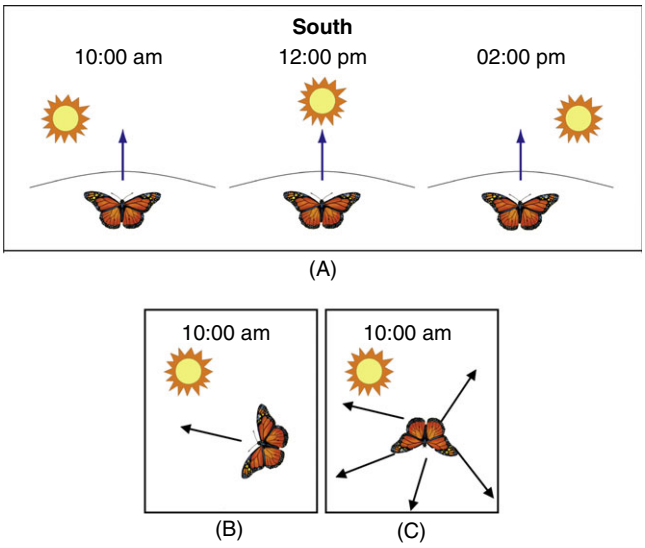


FIGURE B39.6 (A) Monarch butterflies use a time-compensated sun compass to orient south during their fall migration. The butterfly circadian clock allows the butterflies to compensate for the movement of the sun. They are thereby able to maintain a constant bearing in the southerly direction over the course of the day. (B) Clock-shift experiment. When monarch butterflies are housed in a light-dark cycle that is advance by 6 hours, relative to ambient lighting, the butterflies perceive the 10:00 a.m. sun to be a 4:00 p.m. sun and shift their orientation counterclockwise to the southeast. (C) Removal of the time-compensation clocks, which reside in the antennae and not in brain, disrupts the migration south, and the butterflies are not able to successfully travel to their overwintering grounds. The arrows depict disrupted flight directions.

Further Reading

Reppert, S. M., Gegear, R. J., & Merlin, C. (2010). Navigational mechanisms of migrating monarch butterflies. *TINS, 33*, 399–406.

The SCN-controlled circadian rhythm of pineal melatonin production controls photoperiodic responses in mammals. Melatonin is secreted into the bloodstream at night by the pineal gland. The nocturnal duration of melatonin secretion changes in parallel with the length of the night: the melatonin secretory duration thus encodes season. Removal of the pineal gland prevents seasonal responses to changes in day length, while administration of melatonin in patterns that mimic the profiles occurring at different times of year can predictably drive reproduction and other photoperiodic responses.

The actions of melatonin are mediated by specific, G-protein coupled receptors. Two receptors subtypes with high affinity for melatonin have been identified in mammals. The MT1 (Mel_{1a}) receptor subtype is responsible for the majority of melatonin receptor binding found in SCN and elsewhere in brain, while the MT2 (Mel_{1b}) receptor is expressed at considerably lower levels. Several photoperiodic species lack a functional MT2 receptor gene, suggesting that the MT1 receptor may be key for regulating seasonal responses.

A site containing a high concentration of MT1 receptors is the pituitary pars tuberalis (PT). Melatonin acting in the PT affects prolactin release from the anterior pituitary gland. In addition, melatonin acts in the PT to affect gonadotrophin secretion by a circuitous route that has just been unraveled. Melatonin regulates thyroid stimulating hormone (TSH) beta expression in the PT by controlling the expression of Eya3. This transcriptional coactivator interacts with the circadian transcription factor, TEF, which binds to a D-box in the TSHβ promoter, stimulating expression (Dardente et al., 2010). TSH from the PT then acts through TSH receptors located in the mediobasal hypothalamus, altering the balance between synthetic and catabolic enzymes regulating thyroid hormone. A local increase in the biologically active form of thyroid hormone, T_3, within the mediobasal hypothalamus promotes GnRH secretion, leading to gonadotropin secretion and reproduction function. Melatonin, by regulating this local TSH-to-T_3 cascade, has a profound impact on reproductive function in seasonally breeding species.

Summary

The proximate output of the SCN is a rhythm of neuronal firing, which results in rhythmic synaptic and perhaps humoral activation of additional brain regions. The major SCN target is a region dorsal and caudal of the SCN, including the subparaventricular zone (SPZ) and extending into the dorsomedial hypothalamic nucleus (DMH). This extensive projection appears to amplify SCN synaptic output and provides a mechanism for the SCN to influence a wide variety of downstream targets in a manner that can be site-specific. The anatomical pathways leading to specific behavioral and physiological rhythms have been defined to varying extents. Already within the subparaventricular zone, there is divergence of the output pathway controlling rhythms in sleep/wake and locomotor activity (which requires the ventral SPZ) from the pathway regulating body temperature (which requires the dorsal SPZ).

One of the most extensively studied outputs is the pathway regulating melatonin production, involving the SCN, PVN, and sympathetic innervation of the pineal gland. The circadian regulation of melatonin secretory duration regulates reproduction and other photoperiodic responses in seasonally breeding species. The regulation of photoperiodic responses is the most obviously adaptive role for circadian rhythmicity in mammals.

GENERAL SUMMARY

Behavioral and physiological rhythmicity is widespread throughout the animal kingdom. A capacity for cell-autonomous rhythmicity exists in many cell types. Circadian rhythmicity in vertebrates and flies is based on an intracellular molecular feedback loop. Many components of this molecular feedback loop have been elucidated in the past 20 years. There is frequently redundancy of function between closely related genes in mammals. The SCN serve as the coordinating pacemaker in the mammalian circadian timing system. In the absence of the SCN, rhythms in other tissues generally degrade because the synchronization between oscillators is lost. The SCN clock is entrained primarily by light. Lighting information is detected by a specific luminance-coding retinal system that involves intrinsically photosensitive retinal ganglion cells that also serve as a conduit for information originating in rod and cone photoreceptors. SCN axonal output is focused dorso-caudally to the subparaventricular zone, and then diverges along specific anatomical pathways to control a diverse array of rhythmic functions. Among the most physiologically important output rhythms in mammals is the regulation of pineal melatonin, as this hormone is key to regulating photoperiodic adjustments in response to changing day length. Other circadian outputs are of less obvious survival value, but an increasing understanding of the importance of the circadian influence on sleep/wake, metabolism, cardiovascular and neuroendocrine function, and resistance to toxicological challenges indicates that the circadian system dictates an "internal temporal order" that improves physiological and mental performance.

References

Aton, S. J., & Herzog, E. D. (2005). Come together, right … now: Synchronization of rhythms in a mammalian circadian clock. *Neuron, 48,* 531–534.

Aton, S. J., Colwell, C. S., Harmar, A. J., Waschek, J., & Herzog, E. D. (2005). Vasoactive intestinal polypeptide mediates circadian rhythmicity and synchrony in mammalian clock neurons. *Nature Neuroscience, 8,* 476–483.

Atkinson, S. E., Maywood, E. S., Chesham, J. E., Wozny, C., Colwell, C. S., Hastings, M. H., & Williams, S. R. (2011). Cyclic AMP signaling control of action potential firing rate and molecular circadian pacemaking in the suprachiasmatic nucleus. *Journal of Biological Rhythms, 26,* 210–220.

Atwood, A., DeConde, R., Wang, S. S., Mockler, T. C., Sabir, J. S., Ideker, T, & Kay, S. A. (2011). Cell-autonomous circadian clock of hepatocytes drives rhythms in transcription and polyamine synthesis. *Proceedings of the National Academy of Sciences of the United States of America, 108,* 18560–18565.

Balsalobre, A., Damiola, F., & Schibler, U. (1998). A serum shock induces circadian gene expression in mammalian tissue culture cells. *Cell, 93,* 929–937.

Barion, A., & Zee, P. C. (2007). A clinical approach to circadian rhythm sleep disorders. *Sleep Medicine, 8,* 566–577.

Buhr, E. D., Yoo, S. H., & Takahashi, J. S. (2010). Temperature as a universal resetting cue for mammalian circadian oscillators. *Science, 330,* 379–385.

Bunger, M. K., Wilsbacher, L. D., Moran, S. M., Clendenin, C., Radcliffe, L. A., Hogenesch, J. B., … Bradfield, C. A. (2000). *Mop3* is an essential component of the master circadian pacemaker in mammals. *Cell, 103,* 1009–1017.

Cailotto, C., La Fleur, S. E., Van Heijningen, C., Wortel, J., Kalsbeek, A., Feenstra, M, … Buijs, R. M. (2005). The suprachiasmatic nucleus controls the daily variation of plasma glucose via the autonomic output to the liver: Are the clock genes involved? *European Journal of Neuroscience, 22,* 2531–2540.

Chen, S-K., Badea, T. C., & Hattar, S. (2011). Photoentrainment and pupillary light reflexes are mediated by distinct populations of ipRGCs. *Nature, 476,* 92–95.

Chou, T. C., Scammell, T. E., Gooley, J. J., Gaus, S. E., Saper, C. B., & Lu, J. (2003). Critical role of dorsomedial hypothalamic nucleus in a wide range of behavioral circadian rhythms. *Journal of Neuroscience, 23,* 10691–10702.

Czeisler, C. A., Buxton, O. M., & Khalsa, S. B. S. (2005). The humans circadian timning system and sleep-wake regulation. In M. Kryger, T. Roth & W. Dement (Eds.), *Principles and practice of sleep medicine.* 4th ed., (pp. 375–394). New York: WB Saunders.

Dardente, H., Wyse, C. A, Birnie, M. J, Dupré, S. M., Loudon, A. S., Lincoln, G. A., & Hazlerigg, D. G. (2010). A molecular switch for photoperiod responsiveness in mammals. *Current Biology, 20,* 2193–2198.

DeBruyne, J. P. (2008). Oscillating perceptions: The ups and downs of the CLOCK protein in the mouse circadian system. *Journal of Genetics, 87,* 437–446.

DeBruyne, J. P., Weaver, D. R., & Reppert, S. M. (2007). CLOCK and NPAS2 have overlapping roles in the suprachiasmatic circadian clock. *Nature Neuroscience, 10,* 543–545.

Deery, M. J., Maywood, E. S., Chesham, J. E., Sládek, M., Karp, N. A., Green, E. W., … Hastings, M. H. (2009). Proteomic analysis reveals the role of synaptic vesicle cycling in sustaining the suprachiasmatic circadian clock. *Current Biologly, 19,* 2031–2036.

de la Iglesia, H. O., & Schwartz, W. J. (2006). Timely ovulation: Circadian regulation of the female hypothalamo-pituitary-gonadal axis. *Endocrinology, 147,* 1148–1153.

de la Iglesia, H. O., Meyer, J., Carpino, A., Jr, & Schwartz, W. J. (2000). Antiphase oscillation of the left and right suprachiasmatic nuclei. *Science, 290,* 799–801.

de la Iglesia, H. O., Meyer, J., & Schwartz, W. J. (2003). Lateralization of circadian pacemaker output: activation of left- and right-sided luteinizing hormone-releasing hormone neurons involves a neural rather than a humoral pathway. *Journal of Neuroscience, 23,* 7412–7414.

Ecker, J. L., Dumitrescu, O. N., Wong, K. Y., Alam, N. M., Chen, S. K, LeGates, T., … Hattar, S. (2010). Melanopsin-expressing retinal ganglion cell photoreceptors: Cellular diversity and role in pattern vision. *Neuron, 67,* 49–60.

Egli, M., Bertram, R., Sellix, M. T., & Freeman, M. E. (2004). Rhythmic secretion of prolactin in rats: Action of oxytocin coordinated by vasoactive intestinal polypeptide of uprachiasmatic nucleus origin. *Endocrinology, 145,* 3386–3394.

Gatfield, D., & Schibler, U. (2007). Proteasomes keep the circadian clock ticking. *Science, 316,* 1135–1136.

Granados-Fuentes, D., Prolo, L. M., Abraham, U., & Herzog, E. D. (2004). The suprachiasmatic nucleus entrains, but does not sustain, circadian rhythmicity in the olfactory bulb. *Journal of Neuroscience, 24,* 615–619.

Güler, A. D., Ecker, J. L., Lall, G. S., Haq, S., Altimus, C. M., Liao, H. W., … Hattar, S. (2008). Melanopsin cells are the principal conduits for rod-cone input to non-image-forming vision. *Nature, 453,* 102–105.

Hastings, M. H., Field, M. D., Maywood, E. S., Weaver, D. R., & Reppert, S. M. (1999). Differential regulation of mPER1 and mTIM proteins in mouse suprachiasmatic nuclei: new insights into a core clock mechanism. *Journal of Neuroscience, 19*, RC11.

Hatori, M., Le, H., Vollmers, C., Keding, S. R., Tanaka, N., Buch, T., … Panda, S. (2008). Indicuble ablation of melanopsin-expressing retinal ganglion cells reveals their central role in non-image forming visual responses. *PLoS ONE, 3,* e2451.

Hattar, S., Lucas, R. J., Mrosovsky, N., Thompson, S., Douglas, R. H., Hankins, M. W., Yau, KW., et al. (2003). Melanopsin and rod-cone photoreceptive systems account for all major accessory visual functions in mice. *Nature, 424,* 76–81.

Hattar, S., Kumar, M., Park, A., Tong, P., Tung, J., Yau, K. W., & Berson, D. M. (2006). Central projections of melanopsin-expressing retinal ganglion cells in the mouse. *Journal of Comparative Neurology, 497,* 326–349.

Helfrich-Förster, C. (2005). PDF has found its receptor. *Neuron, 48,* 161–163.

Ishida, A., Mutoh, T., Ueyama, T., Bando, H., Masubuchi, S., Nakahara, D., … Okamura, H. (2005). Light activates the adrenal gland: Timing of gene expression and glucocorticoid release. *Cell Metabolism, 2,* 297–307.

Itri, J. N., Michel, S., Vansteensel, M. J., Meijer, J. H., & Colwell, C. S. (2005). Fast delayed rectifier potassium current is required for circadian neural activity. *Nature Neuroscience, 8,* 650–656.

Jin, X., Shearman, L. P., Weaver, D. R., Zylka, M. J., de Vries, G. J., & Reppert, S. M. (1999). A molecular mechanism regulating rhythmic output from the suprachiasmatic circadian clock. *Cell, 96,* 57–68.

Kalsbeek, A., Garidou, M. L., Palm, I. F., Van Der Vliet, J., Simonneaux, V., Pévet, P., & Buijs, R. M. (2000). Melatonin sees the light: Blocking GABA-ergic transmission in the paraventricular nucleus induces daytime secretion of melatonin. *European Journal of Neuroscience, 12,* 3146–3154.

Kalsbeek, A., Ruiter, M., La Fleur, S. E., Cailotto, C., Kreier, F., & Buijs, R. M. (2006). The hypothalamic clock and its control of glucose homeostasis. *Progress in Brain Research, 153,* 283–307.

Kornmann, B., Schaad, O., Bujard, H., Takahashi, J. S., & Schibler, U. (2007). System-driven and oscillator-dependent circadian transcription in mice with a conditionally active liver clock. *PLoS Biol, 5,* e34.

Kudo, T., Loh, D. H., Kuljis, D., Constance, C., & Colwell, C. S. (2011). Fast delayed rectifier potassium current: Critical for input and output of the circadian system. *Journal of Neuroscience, 31,* 2746–2755.

Kuhlman, S. J., & McMahon, D. G. (2006). Encoding the ins and outs of circadian pacemaking. *Journal of Biological Rhythms, 21,* 470–481.

Liu, A. C., Welsh, D. K., Ko, C. H., Tran, H. G., Zhang, E. E., Priest, AA, … Kay, S. A. (2007). Intercellular coupling confers robustness against mutations in the SCN circadian clock network. *Cell, 129,* 605–616.

Lu, J., Zhang, Y. H., Chou, T. C., Gaus, S. E., Elmquist, J. K., Shiromani, P., & Saper, C. B. (2001). Contrasting effects of ibotenate lesions of the paraventricular nucleus and subparaventricular zone on sleep-wake cycle and temperature regulation. *Journal of Neuroscience, 21,* 4864–4874.

Marcheva, B., Ramsey, K. M., Buhr, E. D., Kobayashi, Y., Su, H., Ko, C. H., … Bass, J. (2010). Disruption of the clock components CLOCK and BMAL1 leads to hypoinsulinaemia and diabetes. *Nature, 466,* 627–631.

Maywood, E. S., Reddy, A. B., Wong, G. K., O'Neill, J. S., O'Brien, J. A., McMahon, D. G., … Hastings, M. H. (2006). Synchronization and maintenance of timekeeping in suprachiasmatic circadian clock cells by neuropeptidergic signaling. *Current Biology, 16,* 599–605.

Meredith, A. L., Wiler, S. W., Miller, B. H., Takahashi, J. S., Fodor, A. A., Ruby, N. F., & Aldrich, R. W. (2006). BK calciumactivated potassium channels regulate circadian behavioral rhythms and pacemaker output. *Nature Neuroscience, 9,* 1041–1049, 1193.

Miller, B. H., McDearmon, E. L., Panda, S., Hayes, K. R., Zhang, J., Andrews, J. L., … Takahashi, J. S. (2007). Circadian and CLOCK-controlled regulation of the mouse transcriptome and cell proliferation. *Proceedings of the National Academy of Sciences of the United States of America, 104,* 3342–3347.

Moore, R. Y., & Eichler, V. B. (1972). Loss of a circadian adrenal corticosterone rhythm following suprachiasmatic lesions in the rat. *Brain Research, 42,* 201–206.

Morin, L. P. (2007). SCN organization reconsidered. *J Biological Rhythms, 22,* 3–13.

Nagoshi, E., Saini, C., Bauer, C., Laroche, T., Naef, F., & Schibler, U. (2004). Circadian gene expression in individual fibroblasts: Cell-autonomous and self-sustained oscillators pass time to daughter cells. *Cell, 119,* 693–705.

Pendergast, J. S., Friday, R. C., & Yamazaki, S. (2010). Photic entrainment of *Period* mutant mice is predicted from their phase response curves. *Journal of Neuroscience,, 30,* 12179–12184.

Plautz, J. D., Kaneko, M, Hall, J. C., & Kay, S. A. (1997). Independent photoreceptive circadian clocks throughout *Drosophila*. *Science, 278,* 1632–1635.

Provencio, I. (2011). The hidden organ in our eyes. *Scientific American, 304,* 57–61.

Provencio, I., Jiang, G., De Grip, W. J., Hayes, W. P., & Rollag, M. D. (1998). Melanopsin: An opsin in melanophores, brain, and eye. *Proceedings of the National Academy of Sciences of the United States of America, 95,* 340–345.

Ralph, M. R., Foster, R. G., Davis, F. C., & Menaker, M. (1990). Transplanted suprachiasmatic nucleus determines circadian period. *Science, 247,* 975–978.

Reppert, S. M., & Weaver, D. R. (2002). Coordination of circadian timing in mammals. *Nature, 418,* 935–941.

Sadacca, L. A., Lamia, K. A., deLemos, A. S., Blum, B., & Weitz, C. J. (2011). An intrinsic circadian clock of the pancreas is required for normal insulin release and glucose homeostasis in mice. *Diabetologia, 54,* 120–124.

Saper, C. B., Lu, J., Chou, T. C., & Gooley, J. (2005). The hypothalamic integrator for circadian rhythms. *Trends in Neuroscience, 28,* 152–157.

Schmidt, T. M., Chen, S. K., & Hattar, S. (2011). Intrinsically photosensitive retinal ganglion cells: Many subtypes, diverse functions. *Trends in Neuroscience, 34,* 572–580.

Shearman, L. P., Zylka, M. J., Weaver, D. R., Kolakowski, L. F., Jr, & Reppert, S. M. (1997). Two *Period* homologs: Circadian expression and photic regulation in the suprachiasmatic nuclei. *Neuron, 19,* 1261–1269.

Shi, S., Hida, A., McGuinness, O. P., Wasserman, D. H., Yamazaki, S., & Johnson, C. H. (2010). Circadian clock gene *Bmal1* is not essential; Functional replacement with its paralog, *Bmal2*. *Current Biology, 20,* 316–321.

Silver, R., LeSauter, J., Tresco, P. A., & Lehman, M. N. (1996). A diffusible coupling signal from the transplanted suprachiasmatic nucleus controlling circadian locomotor rhythms. *Nature, 382,* 810–813.

Silver, R., & Schwartz, W. J. (2005). The suprachiasmatic nucleus is a functionally heterogeneous timekeeping organ. *Methods Enzymol, 393,* 451–465.

Stephan, F. K. (2002). The "other" circadian system: Food as a Zeitgeber. *Journal of Biology Rhythms, 17,* 284–292.

Stephan, F. K., & Zucker, I. (1972). Circadian rhythms in drinking behavior and locomotor activity of rats are eliminated by hypothalamic lesions. *Proceedings of the National Academy of Sciences of the United States of America, 69,* 1583–1586.

Tataroglu, O., Davidson, A. J., Benvenuto, L. J., & Menaker, M. (2006). The methamphetamine-sensitive circadian oscillator (MASCO) in mice. *Jouranl of Biology Rhythms, 21,* 185–194.

Ueda, H. R., Hayashi, S., Chen, W., Sano, M., Machida, M., Shigeyoshi, Y, … Hashimoto, S. (2005). System-level identification of transcriptional circuits underlying mammalian circadian clocks. *Nature Genetics, 37,* 187–192.

Vollmers, C., Gill, S., DiTacchio, L., Pulivarthy, S. R., Le, H. D., & Panda, S. (2009). Time of feeding and the intrinsic circadian clock drive rhythms in hepatic gene expression. *Proceedings of the National Academy of Sciences of the United States of America, 106,* 21453–21458.

Weaver, D. R. (1998). The suprachiasmatic nucleus: A 25-year retrospective. *Journal of Biology Rhythms, 13,* 100–112.

Welsh, D. K., Logothetis, D. E., Meister, M., & Reppert, S. M. (1995). Individual neurons dissociated from rat suprachiasmatic nucleus express independently phased circadian firing rhythms. *Neuron, 14,* 697–706.

Welsh, D. K., Yoo, S. H., Liu, A. C., Takahashi, J. S., & Kay, S. A. (2004). Bioluminescence imaging of individual fibroblasts reveals persistent, independently phased circadian rhythms of clock gene expression. *Current Biology, 14,* 2289–2295.

Welsh, D. K., Takahashi, J. S., & Kay, S. A. (2010). Suprachiasmatic nucleus: Cell autonomy and network properties. *Annual Review of Physiology, 72,* 551–577.

Zylka, M. J., Shearman, L. P., Weaver, D. R., & Reppert, S. M. (1998). Three period homologs in mammals: Differential light responses in the suprachiasmatic circadian clock and oscillating transcripts outside of brain. *Neuron, 20,* 1103–1110.

Suggested Readings

Arendt, J. (1995). *Melatonin and the mammalian pineal gland*. London: Chapman & Hall. 1995.

Benca, R, Duncan, M. J., Frank, E., McClung, C., Nelson, R. J., & Vicentic, A. (2009). Biological rhythms, higher brain function, and behavior: Gaps, opportunities, and challenges. *Brain Research Reviews, 62,* 57–70.

Dibner, C., Schibler, U., & Albrecht, U. (2010). The mammalian circadian timing system: Organization and coordination of central and peripheral clocks. *Annual Review of Physiology, 72,* 517–549.

Dunlap, J. C., Loros, J, & DeCoursey, P. J. (Eds.), (2004). *Chronobiology: Biological timekeeping*. Sunderland, MA: Sinauer Associates., 406 pp.

Klein, D. C., Moore, R. Y., & Reppert, S. M. (Eds.), (1991). *Suprachiasmatic nucleus: The mind's clock*. New York: Oxford University Press.

David R. Weaver and Patrick Emery

This page intentionally left blank

CHAPTER

40

The Neurobiology of Sleep and Dreaming

Sleep is a behavioral state that alternates with waking and occupies one-third of our lives. It is characterized by a recumbent posture, a raised threshold to sensory stimulation, a low level of motor output, and a unique behavior: dreaming.

In contrast to sleep, the conscious behavior of waking is characterized by an active and deliberate sensorimotor discourse with the environment. For maintenance of waking behavior, neural gates must remain open for sensory input and motor output, the brain must be tuned and activated, and the chemical microclimate must be appropriate for processing and recording of information.

THE TWO STATES OF SLEEP: RAPID EYE MOVEMENT (REM) AND NON–RAPID EYE MOVEMENT (NREM) SLEEP

Sleep Architecture

The electrical activity of the brain during sleep reveals two distinct substates—rapid eye movement (REM) and non-REM (NREM) sleep—that alternate with a cycle of about 90 minutes (Fig. 40.1B), indicating their control by an endogenous neuronal oscillator. NREM phases are deeper and longer early in the night, whereas the proportion of each cycle occupied by REM is much greater during cycles later in the night (Fig. 40.1A).

NREM and REM sleep are characterized and distinguished using polysomnography (PSG), an electrophysiological recording technique that employs electroencephalography (EEG), electrooculography (EOG), and electromyography (EMG) (Fig. 40.2, middle panel). Combinations of tonic (enduring) and phasic (transient) events in these three types of channels traditionally define sleep stages and substages using the Rechtschaffen and Kales (1968) manual, recently revised with new nomenclature (Iber, Ancoli-Israel, Chesson, & Quan, 2007). The EEG channel tracings of each of these stages are shown in Figure 40.3.

Sleep onset—that is, the onset of Stage 1 NREM sleep (N1 in new nomenclature)—is gradual and characterized by slowing in brain wave frequency from the waking EEG that consists predominantly of beta (14–25 Hz) and gamma (30–80 Hz) frequencies when the eyes are open (Fig. 40.3, first trace) and alpha (8–13 Hz) when the eyes are closed (Fig. 40.3, second trace), to a predominance of theta rhythm (4–7 Hz) in N1 (Fig. 40.3, third trace). During N1, characteristic slow eye movements are recorded in the EOG, EMG muscle tone is diminished relative to waking, awareness of the outside world is lost, and "hypnogogic" dreamlike imagery may occur. As NREM sleep deepens through NREM Stages 2, 3, and 4 (now termed N2 and N3, in which N3 encompasses prior Stages 3 and 4), muscle tone further diminishes, and, in the EEG, changes consist of a progressive decrease in the predominant EEG frequency that is accompanied by a progressive increase in average wave amplitude.

As N1 transitions to N2 (Fig. 40.3, fourth trace), the EEG further slows in the predominantly theta range, and two distinctive phasic wave forms, sleep spindles (De Gennaro & Ferrara, 2003) and K-complexes (Colrain, 2005) occur. Sleep spindles are intermittent clusters of spikes at sigma frequency (12–16 Hz) with characteristic waxing and waning morphology. Spindles occur in two varieties: slow spindles with frequencies approximately 12–14 Hz that appear most strongly in frontal scalp EEG leads and fast spindles (approximately 14–16 Hz) that appear most strongly at centro-parietal EEG leads (Merica, 2000). Sleep spindles are generated by a distinct thalamocortical circuitry detailed in a later section. K-complexes consist of a large negative (upward) deflection in the EEG followed by a positive (downward) deflection and are also generated by thalamo-cortical interactions.

© 2013 Elsevier Inc. All rights reserved.

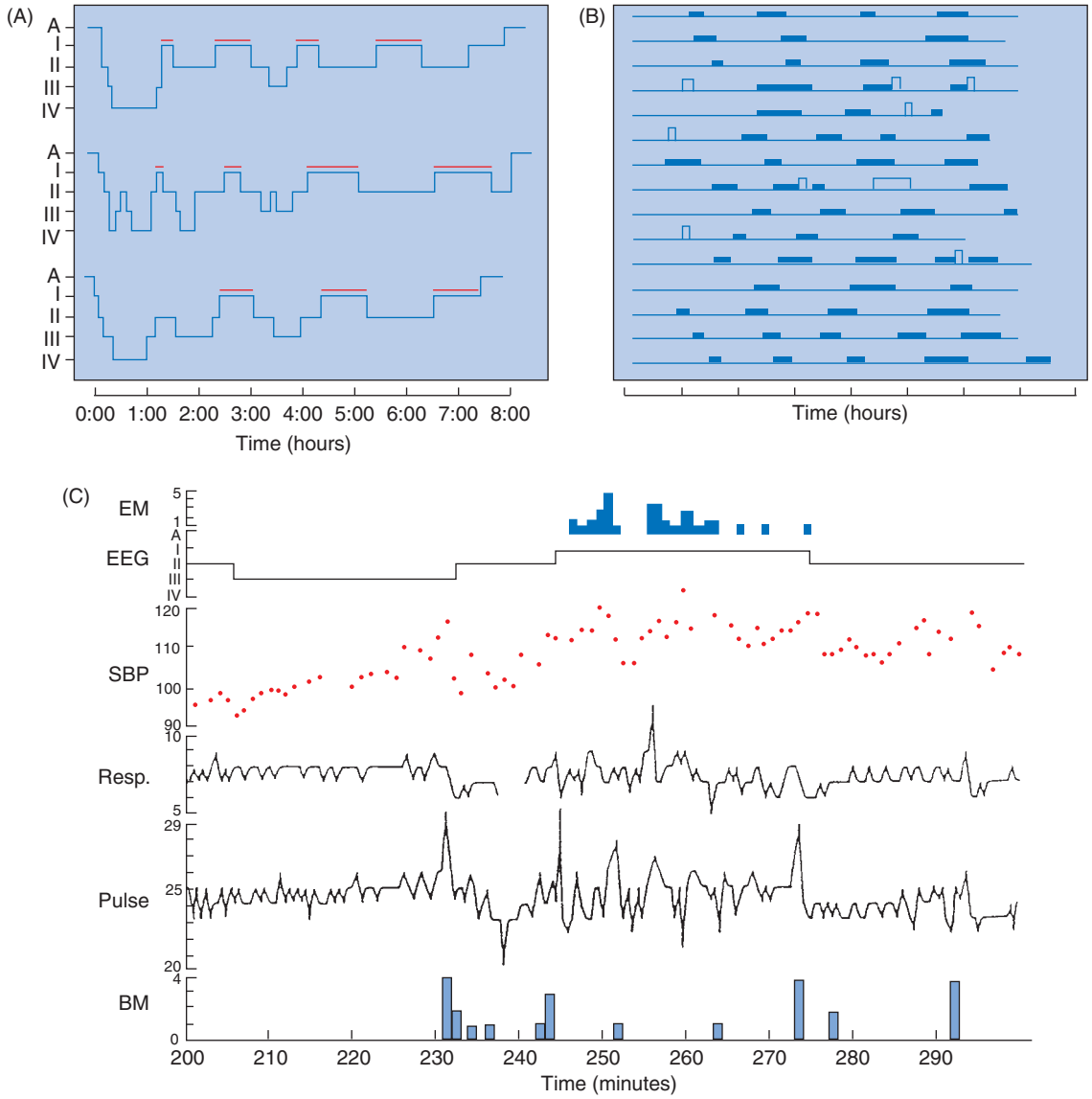


FIGURE 40.1 Periodic activation in sleep cycles. (A) The sleep stages of three people are graphed. The first two or three cycles of the night are dominated by deep stages (3 and 4) of NREM sleep, and REM sleep (indicated by red bars) is brief or nonexistent. During the last two cycles of the night, NREM sleep is lighter (stage 2), and REM episodes are longer, sometimes more than an hour. (B) Fifteen nights of sleep. Each line represents one night of sleep, with REM periods shown as solid bars and periods of wake as taller, open bars. Each record begins at the onset of sleep. The amount of time before the first episode of REM varies, but once REM has begun, the interval between episodes is fairly constant. (C) Eye movements (EM) in numbers per minute, EEG (as scored sleep stage), systolic blood pressure (SBP) in mmHg, respiration (Resp.) in breaths per minute, pulse in beats per quarter minute, and body movement (BM) in numbers per minute over 100 minutes of uninterrupted sleep. The interval from 242 to 273 minutes is considered the REM period, although eye movements are not continuous during that interval.

In N3, also known as slow wave (SWS) or delta sleep (Fig. 40.3, fifth trace), many fewer sleep spindles occur, and EEG frequency further slows and displays increasing predominance of high amplitude delta (0.5–4.0 Hz) waves (Dijk, 2009). In the early 1990s, a yet slower oscillatory rhythm of less than 1 Hz, the "slow oscillation," was described in cats and, shortly afterward, in humans (Steriade, 2006). The slow oscillation shows maximal spectral power in SWS and exerts an organizing or "grouping effect" on delta oscillations, spindles, and K-complexes (Steriade, 2006). As NREM deepens from N1 to N3, the threshold for arousal rises in proportion to the degree of EEG slowing, and during N3, awakenings are difficult, incomplete and brief. Those awakened from this state are confused, have difficulty reporting conscious experience, and return to sleep rapidly, especially early in the night. While dreaming also occurs during NREM sleep (Fig. 40.2, bottom panel), it is much more strongly associated with the activated brain state of REM.

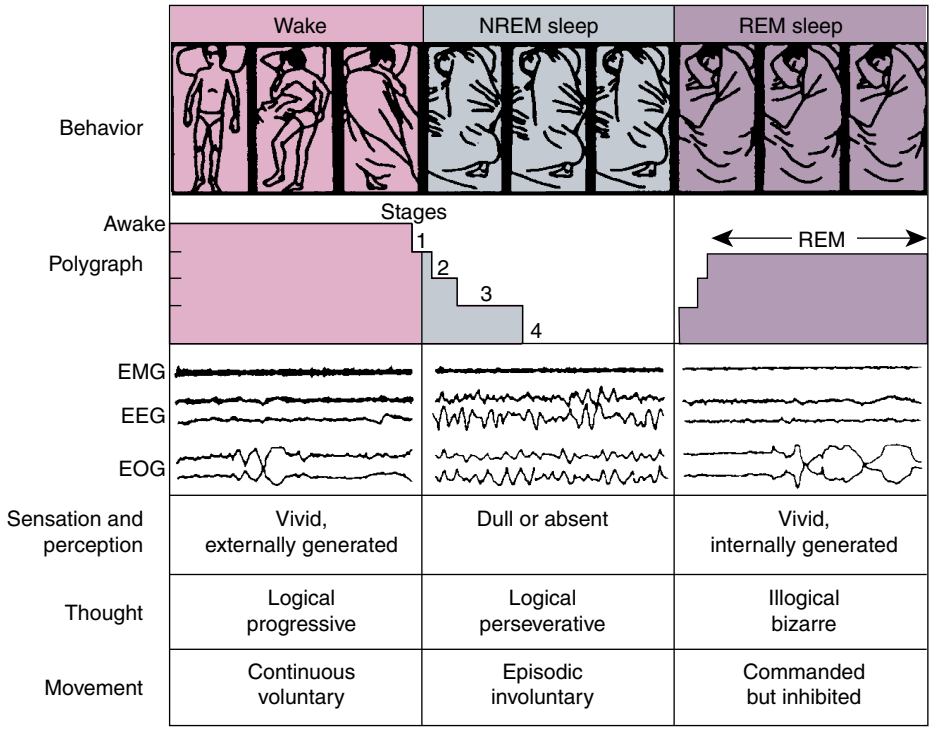


FIGURE 40.2 Behavioral states in humans. Body position changes during waking and at the time of phase changes in the sleep cycle. Removal of facilitation (during stages 1–4 of NREM sleep) and addition of inhibition (during REM sleep) account for immobility during sleep. In dreams, we imagine that we move, but no movement occurs. Tracings of electrical activity are shown in ~20-s sample records. The amplitude of the electromyogram (EMG) is highest in waking, intermediate in NREM sleep, and lowest in REM sleep. The electroencephalogram (EEG) and electrooculogram (EOG) are activated in waking and REM sleep and inactivated in NREM sleep. *Reprinted with permission from Hobson and Steriade (1986).*

REM sleep, discovered in 1953 by Eugene Aserinsky and Nathaniel Kleitman (Aserinsky & Kleitman, 1953), has also been termed "paradoxical" sleep because, whereas its fast mixed-frequency EEG resembles waking (Fig. 40.3, sixth trace), muscle tone reaches its minimum level (see Mallick et al., 2011, for extensive reviews). This EMG minimum in REM results from central, brainstem-mediated inhibition of spinal alpha motor neurons ("REM atonia"; Chase, 2008). Postural shifts precede and follow REM sleep (Fig. 40.2, upper panel). REM has also been termed "active" sleep because of its wakelike low-amplitude, high-mixed-frequency EEG, as well as "desynchronized" sleep, in contrast with "synchronized" NREM, especially SWS, in which high-amplitude waves reflect synchronous activity of thalamic and cortical neurons (described below). A characteristic EEG feature of REM sleep in humans is the "sawtooth wave" (Fig. 40.3, sixth trace). During REM, unlike waking, there is inhibition of sensory input and motor output. Subjects aroused from REM sleep, especially during periods with frequent eye movements, often give detailed reports of dreams characterized by vivid hallucinations, bizarre thinking, and intense emotion (Fig. 40.2 bottom panel).

As its name suggests, saccadic (rapid) eye movements (REMs) occur during REM (Fig. 40.1C), singly and in clusters, reflecting the fact that, unlike skeletal muscles, oculomotor muscles maintain tonus in REM. REMs are believed to reflect underlying, central phasic neuronal events also appearing as peripheral muscle twitches (that briefly overcome atonia), middle-ear muscle movements (MEMAs), and peri-orbital potentials (PIPs). In humans, phasic REM events are likely related to ponto-geniculo-occipital (PGO) waves. PGO waves are phasic field potentials that, in the REM sleep of cats, are recorded sequentially in the pons (P), lateral geniculate body of the thalamus (G), and the occipital cortex (O) (Datta & Maclean, 2007), and which also have recently been demonstrated in humans (Lim et al., 2007). In contrast to such phasic events, the tonic wakelike REM EEG reflects cortical activation by the cholinergic ascending arousal systems described below (Saper, Fuller, Pederson, Lu, & Scammell, 2010).

Central Autonomic Control Systems Are State Dependent

Sympathetic outflow diminishes in the transition from waking to NREM, and during SWS, parasympathetic outflow is predominant. Many autonomic and regulatory functions, such as heart rate, blood pressure, and respiration rate, diminish in NREM sleep (Fig. 40.1C), but some neuroendocrine activity, such as release of growth hormone from the pituitary, increases (Pannain & Van Cauter, 2007). With the onset of REM, sympathetic influence increases, although not to the level of waking, and cardiovascular and respiratory measures become more variable (Trinder et al., 2001; Fig. 40.1C). Active temperature and respiratory control are greatly diminished in REM sleep (Hobson, 1989). In mammals, sleep and thermoregulation can exert reciprocal influence on one another through shared circuitry in the preoptic anterior hypothalamus (Parmeggiani, 2003).

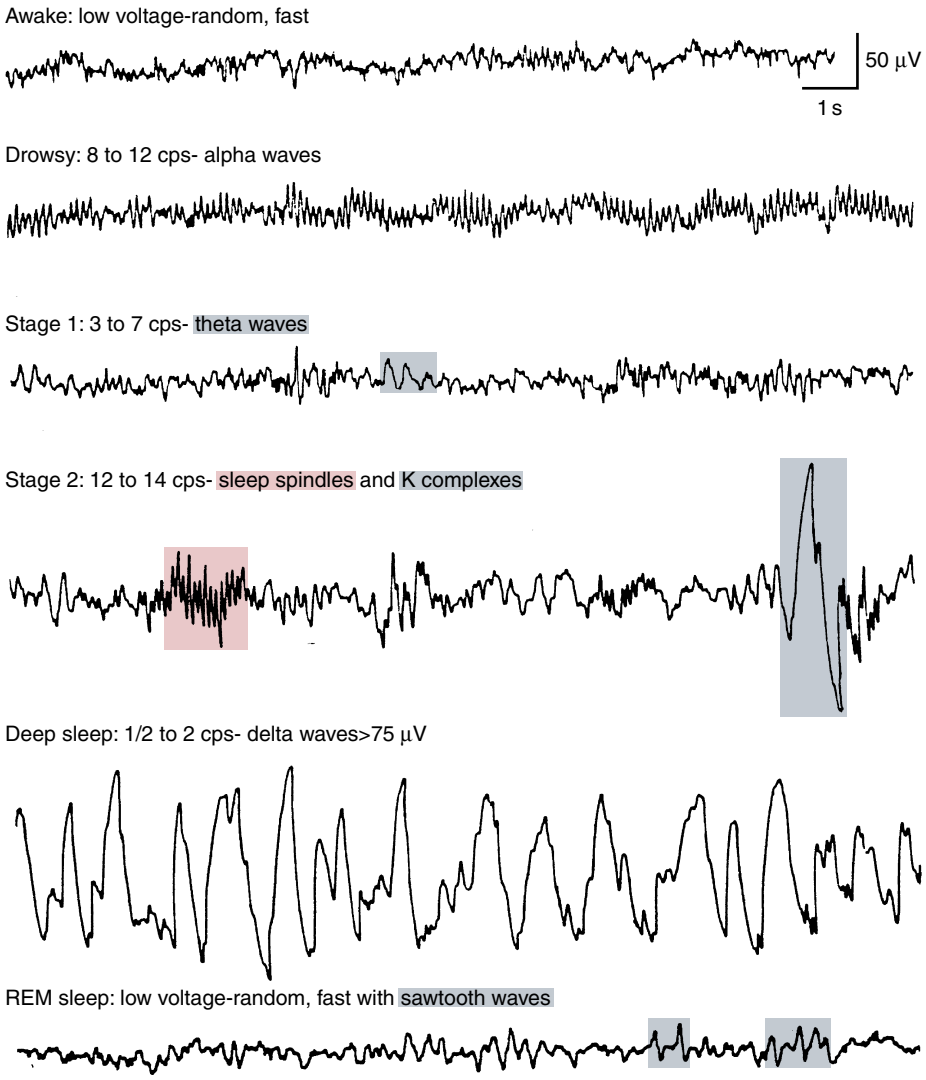

Awake: low voltage-random, fast
50 μV
1 s
Drowsy: 8 to 12 cps- alpha waves
Stage 1: 3 to 7 cps- theta waves
Stage 2: 12 to 14 cps- sleep spindles and K complexes
Deep sleep: 1/2 to 2 cps- delta waves>75 μV
REM sleep: low voltage-random, fast with sawtooth waves

FIGURE 40.3 Electroencephalograms showing electrical activity of the human brain during different stages of sleep.

NEURAL CONTROL OF THE SLEEP–WAKE AND REM/NREM CYCLES

ARAS Sensorimotor and Neuromodulatory Neurons

Most of the classic reticular core neurons have very specific afferent inputs and highly organized outputs that underlie the multiple functions of the brainstem, such as sensory input, motor output, and autonomic control. These reticular neurons receive inputs from skin, muscle, bone, and joint receptors in the periphery, which they integrate and link to the vestibular and cerebellar circuits that determine posture and movement (see Chapter 29). Performing an entirely different function are small groups of neurons that send widely branching axons to distant parts of the brain where their neurotransmitters modulate brain function and play an essential role in behavioral state regulation including control of sleep/wake and REM/NREM cycles. These ARAS modulatory neurons differ from neighboring sensorimotor neurons in both morphology and characteristic firing patterns.

Sensorimotor neurons, approximately 50 to 75 microns in diameter, can fire continuously at high rates of up to 50 Hz and can generate bursts of up to 500 Hz. Their larger axons, especially those projecting to the spinal cord, have conduction velocities in excess of 100 m/s (Hobson & Steriade, 1986), making these neurons well suited to rapid posture adjustment and motor control. In contrast, modulatory neurons are smaller (10–25 microns in diameter), and even those with long axons conduct their signals very slowly (1 m/s). Modulatory neurons fire wider spikes (2 ms in duration) at much slower tonic frequencies (1–10 Hz) than sensorimotor neurons. In addition, they

often show a very regular, metronome-like firing pattern, a reflection of their pacemaker properties. As their leaky membranes spontaneously and slowly depolarize, they reach threshold, fire, and then self-inhibit, becoming refractory even to exogenous excitatory inputs (Ramirez, Tryba, & Peña, 2004). Because of their vast postsynaptic domain, they can help set behavioral and mental states.

ARAS Neuronal Groups Involved in Behavioral State Regulation Produce Specific Neuromodulators

Modulatory ARAS neuronal groups produce specific types of neuromodulators (Dahlstrom & Fuxe, 1964). Norepinephrine (NE) neurons are located in the pons and medulla in two major groups. The lateral tegmental group consists of scattered neurons in the ventral and lateral reticular formation and appear to be involved in hypothalamic regulation and motor control (Moore & Card, 1984). Neurons of the second and major NE cell group, the locus coeruleus (LC), are located in the rostral pontine reticular formation and central gray matter. One group of LC neurons projects caudally to sensory regions of the brainstem and spinal cord, and another projects widely to the cerebellar cortex, dorsal thalamus, and cerebral cortex (Moore & Card, 1984). Thus, LC neurons are likely to be involved in regulating sensory input and cortical activation. Serotonin (5-HT) neurons are located in the raphe nuclei that extend, in the ARAS midline, from caudal medulla to midbrain in the dorsal (DR) and median raphe nuclei. These neurons project rostrally and innervate nearly the entire forebrain, again suggesting a role in regulation of behavioral state. The other important set of brainstem reticular formation nuclei involved in behavioral state control produce acetylcholine (ACh) and include two pontine nuclei: the laterodorsal tegmental (LDT) and the pedunculopontine (PPT) nuclei. Cholinergic neurons in these nuclei project to the brainstem reticular formation, hypothalamus, thalamus, and basal forebrain. A second set of cholinergic neurons in the basal forebrain (described below) projects to the limbic forebrain and neocortex. In contrast to these modulatory (5-HT, NE, and ACh) neurons, neurons in brainstem reticular formation nuclei involved in sensorimotor integration typically produce the excitatory transmitter glutamate or the inhibitory amino acids gamma-aminobutyric acid (GABA) and/or glycine.

Waking Requires Active Maintenance

The reticular formation is necessary to maintain the waking conscious state (Moruzzi & Magoun, 1949). Lesions in the midbrain reticular formation result in a state that resembles NREM sleep. Maintenance of the waking state requires brain mechanisms that compete with other active mechanisms that promote sleep. Multiple neurochemical systems maintain cortical arousal, and this redundancy allows the waking state to persist in the face of deficiency or destruction of specific arousal-related neurons (Saper et al., 2010).

These nuclei and their respective neurotransmitters and targets (Fig. 40.4) include the basal forebrain (BF) nuclei (medial septum, nucleus of the diagonal band, substantia innominata–nucleus basalis) that send ACh projections to topographically specific cortical and subcortical limbic areas (Jones, 2008); hypothalamic neurons (i) producing the neuropeptide orexin (also called hypocretin) in the lateral (LH) and perifornical (PFH) hypothalamus that project to other arousal systems as well as the entire cerebral cortex (Ohno & Sakurai, 2008), and (ii) histamine (HA)-producing neurons from the tuberomammilary

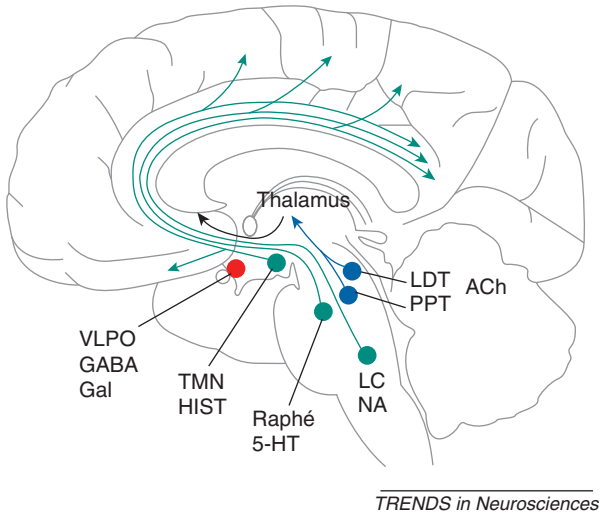


FIGURE 40.4 Ascending arousal systems of the brainstem and posterior hypothalamus and sleep-promoting areas of the anterior hypothalamus. Nuclei producing wake-promoting monoaminergic neuromodulators are shown in green. These include noradrenaline (NA)-producing neurons of the locus coeruleus (LC) and serotonin (5-HT)-producing neurons of the dorsal raphe (Raphe), both located in the mesopontine brainstem, as well as histamine (HIST)-producing neurons of the tuberomammilary nucleus (TMN), located in the posterior hypothalamus. These nuclei project to and activate widely distributed terminal fields throughout the forebrain. Nuclei shown in blue produce acetylcholine (ACh), which promotes forebrain activation in both waking and REM sleep. Cholinergic nuclei include the pedunculopontine (PPT) and laterodorsal (LDT) tegmental nuclei of the mesopontine brainstem as well as cholinergic nuclei of the basal forebrain (not shown). Cholinergic neurons of the PPT and LDT project to the thalamus and basal forebrain that then, in turn, send arousal-promoting projections to both distributed and regionally specific terminal fields in the forebrain. Shown in red are sleep-promoting neurons of the ventrolateral preoptic (VLPO) area of the anterior hypothalamus. VLPO neurons produce the inhibitory neuromodulator gamma-amino butyric acid (GABA) and the inhibitory neuropeptide galanin (Gal). VLPO neurons project to and inhibit the above wake-promoting neurons, thereby promoting and maintaining sleep. In turn, the wake-promoting neurons of the LC, DRN, and TMN reciprocally innervate and inhibit the neurons of the VLPO. *From Saper et al. (2001).*

nucleus (TMN) of the posterior hypothalamus (PH), which send widespread projections to diverse forebrain sites (Szymusiak & McGinty, 2008); midbrain neurons producing dopamine (DA) from specifically wake-active centers in the midbrain periaqueductal gray (PAG; Saper et al., 2010) that are distinct from midbrain substantia nigra pars compacta (SNpc) and ventral tegmental (VTA) area dopaminergic neurons (that, respectively, promote movement via projection to the basal ganglia, and appetitive behaviors via projection to the ventral striatum and prefrontal cortex); and mesopontine (junction of pons and midbrain) nuclei that include (i) ACh-producing PPT and LDT projecting to the diencephalon, basal forebrain, and restricted portions of the cortex; (ii) NE-producing LC neurons that send widespread projections to the forebrain; and (ii) 5-HT-producing DR neurons that also widely project to the forebrain (for reviews, see Pace-Schott & Hobson, 2002; Saper et al., 2010).

Certain neurochemical systems also promote cortical arousal indirectly. For example, cholinergic neurons of the LDT and PPT promote cortical activation and desynchronization of the EEG by exciting thalamo-cortical projection and, especially, inter-laminar neurons of the thalamus, as well as by inhibiting the inhibitory thalamic reticular neurons and exciting cholinergic neurons of the basal forebrain (Steriade, 2000). Similarly, orexinergic neurons stabilize the waking state by exciting other arousal systems in the hypothalamus, basal forebrain, and brainstem (Jones, 2008). Additionally, GABAergic neurons of the basal forebrain promote cortical activation by inhibiting inhibitory GABAergic cortical interneurons (Saper et al., 2010). Notably, at localized sites in the mesopontine brainstem, higher levels of GABA may actually promote wakefulness at the expense of REM and NREM sleep due to similar complex circuits of inhibition and disinhibition (Watson, Baghdoyan, & Lydic, 2011).

The Hypothalamus Controls the Sleep–Wake Transition

Lesions to anterior portions of the hypothalamus cause insomnia, whereas lesions of the posterior hypothalamus cause hypersomnolence (Saper, Chou, & Scammell, 2001). Neurons in the anterior hypothalamus increase firing during sleep, and their electrical stimulation promotes sleep, whereas neurons of posterior portions of the hypothalamus decrease firing during sleep, and their electrical stimulation promotes waking (Szymusiak & McGinty, 2008). Sherin, Shiromani, McCarley, and Saper (1996) demonstrated that neurons in the ventrolateral preoptic area (VLPO) of the anterior hypothalamus increase metabolic activity during NREM sleep in the rat. VLPO neurons produce both GABA and the inhibitory neuropeptide galanin and project to wake-promoting neurons in the TMN, LC, DR, LDT, and PPT (Saper et al., 2001). Activity in an adjacent region, the median preoptic area (MnPO), may respond to increased homeostatic sleep pressure (see next section) and contribute to subsequent activation of the VLPO that then persists to maintain sleep itself (Saper et al., 2010). Additionally, distinct subregions of the VLPO may be specialized for the maintenance of REM and NREM sleep (Saper et al., 2010). The LC, DRN, and TMN, in turn, innervate and inhibit the neurons of the VLPO (Saper et al., 2001).

Saper et al. (2001) suggested that this mutually inhibitory arrangement constitutes the biological analog of a bistable electrical "flip-flop" switch in which either the sleep (VLPO dominant) or wake (LC, DRN, TMN dominant) conditions are self-reinforcing stable states, whereas intermediate states are highly transient (Fig. 40.5). Orexin plays a key modulatory role in this flip-flop mechanism by providing excitatory drive on aminergic neurons, thereby further stabilizing the waking pole of the bipolar sleep–wake switch. In 1999, two research teams independently showed that deficiencies in orexinergic neurotransmission underlie the sleep disorder narcolepsy. Orexin-producing neurons are lost in human narcolepsy (Nishino, Ripley, Overeem, Lammers, & Mignot, 2000), likely due to an autoimmune mechanism, and, as a result, wake to sleep transitions can occur at inappropriate times (Saper et al., 2001). Orexin also opposes the atonia of REM, and, therefore, in narcolepsy, REM atonia can intrude into waking in the form of cataplexy (Alam, Szymusiak, & McGinty, 2011).

Homeostatic and Circadian Factors Interact to Control Sleep Onset and Offset

The timing of transitions between sleep and waking that are effected in the hypothalamus, however, depend upon input from extra-hypothalamic sources. Alexander Borbely's two-process model of sleep propensity (Fig. 40.6) suggests that homeostatic mechanisms (Process S) interact with circadian factors (Process C) to regulate behaviorally expressed sleep and waking (Borbely, 1982). Process C (Fig. 40.6A) constitutes a circadian rhythm (see Chapter 39) in the propensity to fall asleep that is controlled by an endogenous circadian clock in the suprachiasmatic nucleus (SCN) of the anterior hypothalamus (Dibner, Schibler, & Albrecht, 2010). Reliable output from the SCN, possessing a near–exact 24-hour periodicity in humans (Czeisler et al., 1999), results in a daily rhythm in sleep propensity that is maximal in the early morning, corresponding to the circadian nadir in body temperature and is minimal in the evening, corresponding to a "forbidden zone" for sleep (Lavie, 1986). Another smaller increase in circadian sleep propensity, the "post-prandial dip," may occur in the early afternoon around "siesta time"

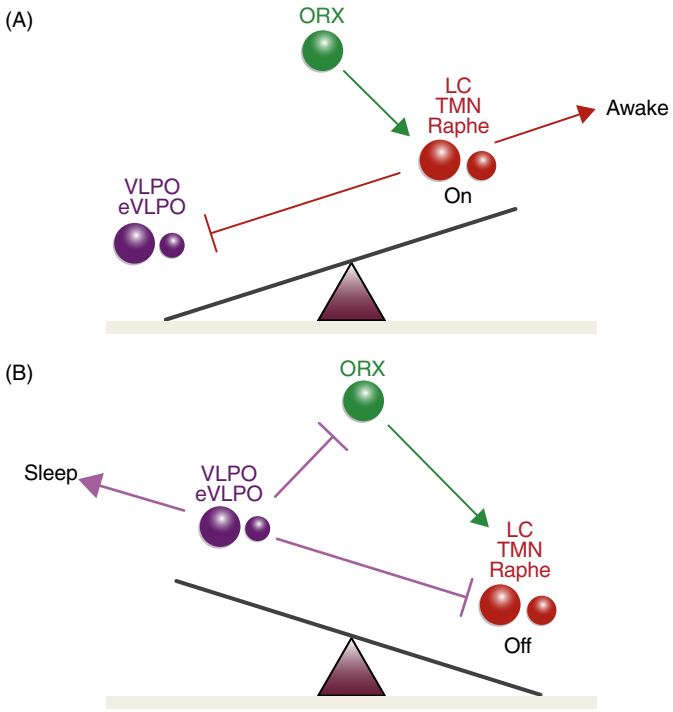


FIGURE 40.5 The flip-flop switch model of sleep–wake control proposed by Saper et al. (2001). (A) During waking, monoaminergic neurons (shown in red) of the locus coeruleus (LC), tuberomammillary (TMN), and dorsal raphe (Raphe) nuclei promote waking and inhibit the sleep-producing neurons of the ventrolateral preoptic area (VLPO) that are shown in purple. Orexinergic neurons (ORX) from the lateral hypothalamus (shown in green) excite and reinforce the wake-promoting activity of the monoaminergic neurons. (B) During sleep, neurons of the VLPO inhibit the wake-promoting monoaminergic neurons as well as the orexinergic neurons that help maintain their activity. By inhibiting the VLPO, monoaminergic neurons disinhibit their own wake-promoting activity, whereas, by inhibiting the monoaminergic nuclei, the VLPO disinhibits its own sleep-promoting activity. This mutually inhibitory interaction produces conditions analogous to an electrical flip-flop switch whereby the alternate states of sleep and waking are self-reinforcing but intermediate states are unstable and transient. This results in rapid transitions from waking to sleep and vice versa. Orexin serves to stabilize the waking state and prevent inappropriate transitions to sleep should the monoaminergic drive transiently weaken. In narcolepsy, orexinergic neurons are lost, thereby allowing such abrupt transitions from waking to sleep during the characteristic sleep attacks of this disorder. (*Modified from Saper, Scammel & Lu, 2005*).

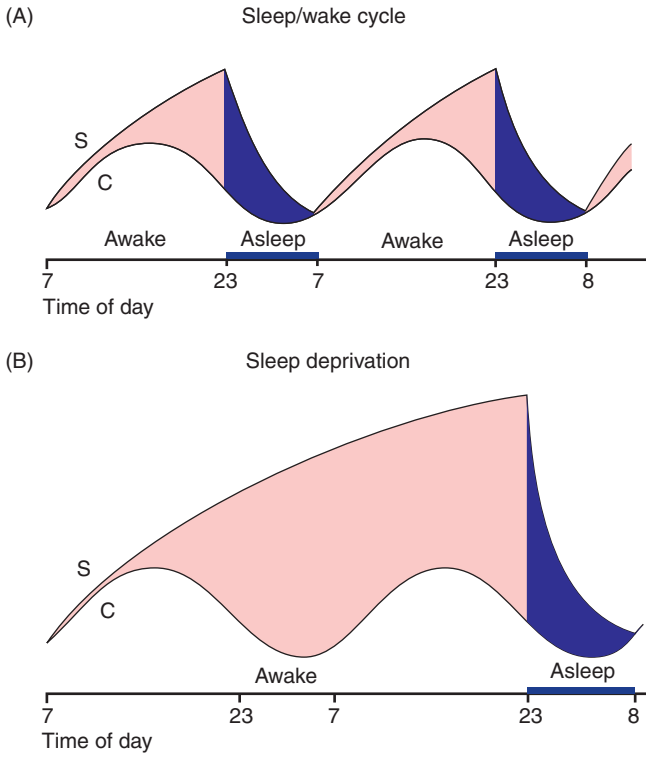


FIGURE 40.6 The Borbely and Daan model of sleep regulation. Sleep is assumed to result from the actions of process C and process S. Process C follows a circadian rhythm and is independent of sleeping and waking. Process S, on the other hand, depends on sleep–wake behavior; S declines during sleep and rises continuously during sleep deprivation. The period of recovery sleep that follows sleep deprivation is more intensive but only slightly longer than normal. If curve C represents the threshold for waking up, then at any time, "sleep pressure" is the (vertical) distance between the S and C curves. The greater the distance, the greater the pressure to fall asleep. *Reprinted with permission from Daan et al. (1984).*

(Bes, Jobert, & Schulz, 2009). If a research subject is deprived of sleep, sleepiness will continue to follow a circadian rhythm, indicating that sleepiness cycles can be independent of actual sleeping and waking.

Briefly, the molecular and cellular basis of Process C involves synchronized molecular clocks in each cell of the SCN that maintain 24-hour periodicity via interlocking positive and negative feedback control in the transcription and translation of circadian genes (see Chapter 39 and Dibner et al., 2010). This endogenous clock is entrained to the ambient photoperiod via nonvisual, photoreceptive pigments such as melanopsin in retinal ganglion cells (Peirson & Foster, 2006) that send glutamatergic signals to the SCN via the retinohypothalamic tract. The SCN conveys circadian information via the subparaventricular and dorsomedial nuclei of the hypothalamus to the VLPO to influence sleep–wake transitions (Saper, Chou, & Scammel, 2005). Secretion of the sleep hormone melatonin from the pineal gland is also indirectly controlled by the SCN, increases just prior to normal sleep onset, and provides an internal indicator of circadian time as well as feedback influence on the SCN, serving to further stabilize circadian rhythms via action at melatonin receptors on the SCN itself (Arendt, 2006).

Sleep homeostasis, or Process S, is a mechanism that increases sleep pressure as a function of the amount of time since the last sleep episode (Fig. 40.6B). Process S generates the powerfully increased drive to sleep that accompanies prolonged sleep deprivation and is believed to result from sleep-promoting substances, "endogenous somnogens," that accumulate during waking and dissipate during sleep. The purine nucleoside adenosine,

which accumulates in extracellular space as a product of metabolism, is currently the most likely candidate (Porkka-Heiskanen & Kalinchuk, 2011). Evidence also exists for endogenous somnogenic actions of interleukin-1B, prostaglandin D2, and growth hormone–releasing hormone (Szymusiak & McGinty, 2008). A primary mechanism for the sleep-promoting effects of adenosine is believed to be activation of inhibitory adenosine A1 receptors on wake-promoting cholinergic cells in the basal forebrain that project to and activate the cortex and desynchronize the EEG. However, recent findings suggest that adenosine may also produce somnogenic effects elsewhere in the brain, such as at excitatory A2A receptors in the hypothalamus and ventral striatal areas. Interestingly, the most commonly employed drug to counteract high homeostatic sleep pressure, the nonspecific adenosine receptor agonist caffeine, acts not at the basal forebrain but at postsynaptic A2A receptors in the basal ganglia (Lazarus et al., 2011). Adenosine may also mediate localized sleeplike electrophysiology in restricted areas of the cortex reflecting a prior-use dependent, restorative function (Krueger et al., 2008). Astrocytic release of adenosine, which inhibits local neurons by binding to A1 receptors, may allow localized emergence of sleeplike activity such as the slow (<1 Hz) oscillation (Halassa, 2011).

By acting in concert, processes C and S promote a single, daily consolidated nocturnal sleep bout in healthy adult humans (Dijk & Czeisler, 1994; Fig. 40.6). Specifically, in the evening, when sleep homeostatic pressure is maximal following a day's waking, the circadian "forbidden zone" maintains wakefulness until the normal time of sleep onset. In the early morning at the circadian nadir, when homeostatic sleep pressure has largely dissipated, Process C maintains the drive to sleep until the normal time of awakening.

NREM Sleep Rhythms Require Thalamo-Cortical Interaction

Following the transition from waking to NREM, neuronal firing rates in all wake-active nuclei decline with deepening NREM. This decline allows the emergence of intrinsic oscillations in reciprocally interconnected thalamic and cortical circuits that appear in the human EEG as the characteristic wave forms of N2 and N3 sleep: spindles, K-complexes, delta waves, and the (<1 Hz) slow oscillation (Steriade, 2006; Fig. 40.7). During waking, ascending arousal systems modulate thalamo-cortical circuits with the arousal-promoting neuromodulators (ACh, NE, 5-HT, and HA) and prevent the emergence of these oscillations. Similarly, during REM sleep, thalamo-cortical rhythms are prevented by cholinergic activation of the thalamus from the LDT and PPT.

The sleep spindles, K-complexes, and delta waves of NREM reflect the influence of the cortical slow oscillation on the thalamic reticular nucleus, a thin sheet of GABAergic neurons surrounding the thalamic periphery that inhibit other thalamic neurons (Steriade, 2000). The slow oscillation results from an alternation between a prolonged hyperpolarized ("down-state") phase and a shorter, rapidly

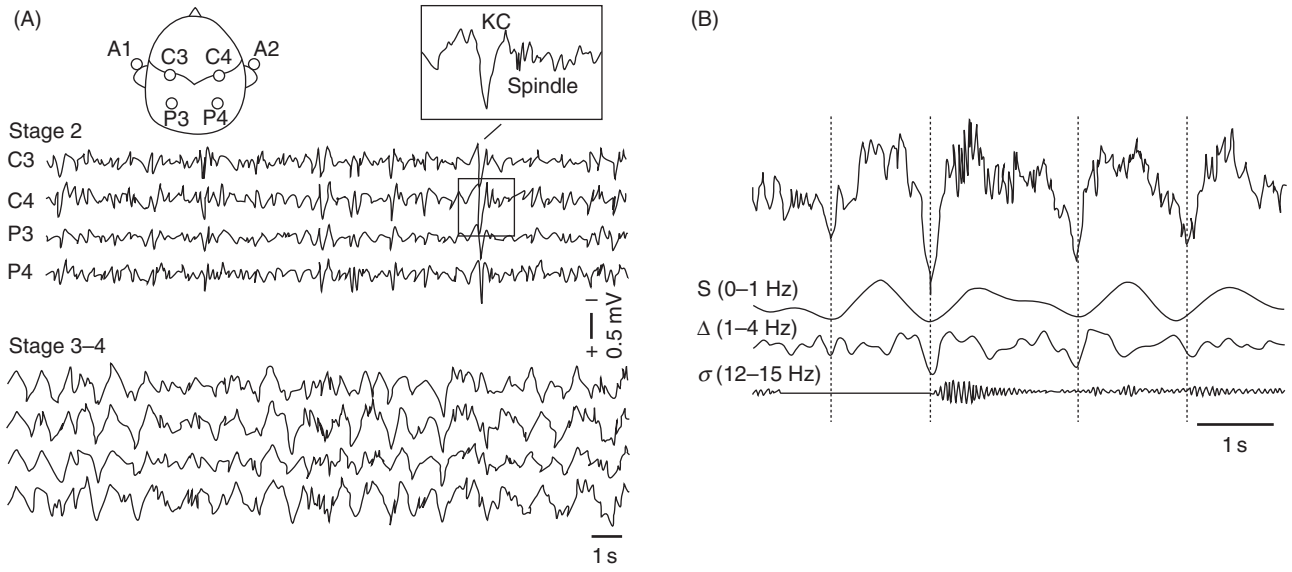


FIGURE 40.7 Characteristic NREM oscillations in the human EEG. (A) Recordings from central (C3, C4) and parietal (P3, P4) scalp sites referenced to the contralateral ears (A1, A2) during stage 2 NREM sleep (upper traces) and stages 3 and 4 NREM or slow wave sleep (lower traces). The insert shows the characteristic Stage 2 K-complex and sleep spindle wave forms. (B) Frequency composition of the NREM EEG showing the slow (0–1 Hz), delta (1–4 Hz), and 12–15 Hz sleep spindle (sigma frequency) oscillations. Note the organizing or grouping effect of the slow oscillation on the faster NREM oscillations. *From Amzica and Steriade (1996).*

spiking, depolarized ("up-state") phase that is synchronized among large assemblages of cortical neurons (Menicucci et al., 2009). The slow oscillation can be generated entirely within the cortex and may be an intrinsic property of cortical networks (Halassa, 2011), although the thalamus may also be crucial to its expression in vivo (Crunelli & Hughes, 2010). In human sleep, slow oscillations originate in frontal areas and travel posteriorly within the cortex (Massimini, Huber, Ferrarelli, Hill, & Tononi, 2004) but can also occur locally in the cortex (Nir et al., 2011). The slow oscillation exerts an organizing or grouping influence on other intrinsic NREM sleep rhythms (Fig. 40.6).

The cortical slow oscillation initiates sleep spindles when, during its depolarized up-phase, neurons projecting from the cortex to the thalamus excite neurons of the thalamic reticular nucleus that, in turn, inhibit thalamo-cortical neurons projecting from the thalamus to the cortex (Steriade, 2000). This periodic inhibition of thalamo-cortical neurons changes their firing pattern from a "transmission mode," in which firing frequency is proportional to the strength of a sensory input, to a "bursting mode," in which transmission of sensory information to the cortex is blocked (McCormick & Bal, 1997; Fig 40.8). This bursting mode is initiated when thalamo-cortical neurons are sufficiently hyperpolarized to activate a unique ion channel that opens only when their membrane potential falls well below its normal resting potential (McCormick & Bal, 1997). Opening of this channel begins a series of membrane events ("H" currents followed by "T" or "low-threshold calcium spike" currents) that culminates in thalamo-cortical neurons emitting a series of spikes that impinge on the cortex at spindle frequency.

Delta frequency EEG oscillations may have both thalamic and cortico-thalamic origins (Steriade, 2000). Delta-frequency oscillations can arise via thalamic neurons' intrinsic delta-frequency pacemaker property or, alternatively, when cortico-thalamic neurons excite thalamic reticular neurons that, in turn, hyperpolarize the thalamo-cortical neurons, causing them to stimulate the cortex at delta frequency. Both of these mechanisms also result from H and T currents. Because spectral power in the delta (Ferrara et al., 2002) as well as the slow oscillation (Bersagliere & Achermann, 2010) frequencies increase in response to the duration of prior waking, both are indicative of increased sleep pressure.

REM Sleep Is Initiated in the Brainstem

In the transition from NREM to REM sleep, firing rates of LC, DR, and TMN neurons and respective levels of NE, 5-HT, and HA at their terminal fields decline to their minimum levels. In contrast, firing rates of PPT/LDT neurons return to levels comparable to waking (reviewed in Pace-Schott & Hobson, 2002). Like aminergic neurons, orexin neurons of the LH/PFH decline in activity at sleep onset and remain quiescent in REM (Ohno & Sakurai, 2008). In the transition from NREM to REM, activation of the thalamus by brainstem cholinergic (LDT and PPT) neurons blocks the intrinsic oscillations of NREM and results in cortical EEG desynchronization (Steriade, 2000). The PPT and LDT also activate cortically projecting BF cholinergic neurons, resulting in further desynchronization of the cortex (Jones, 2008).

The cellular and molecular events that produce the regular alternation of REM and NREM sleep emerge under brainstem control once sleep has been achieved (reviewed in Steriade & McCarley, 2005 and Datta & Maclean, 2007). The cardinal neural features of REM are initiated by specific brainstem cell groups, illustrated in Figure 40.9, and include: (i) EEG desynchronization resulting from ARAS activation of the thalamus and basal forebrain; (ii) PGO waves in the cat, or P-waves in the rat, originating from different distinct pontine generators. PGO waves result from tonic disinhibition and phasic excitation of bursting cells in the peribrachial region of the mesopontine tegmentum and, like REM sleep, they can be triggered and augmented by cholinergic stimulation see below; (iii) rapid eye movements brought about by pre-motor neurons in proximity to oculomotor (abducens) nuclei; (iv) a change in hippocampal EEG from irregular rhythms to a regular theta rhythm (Vertes & Kocsis, 1997); and (v) skeletal muscle atonia resulting when pontine REM-active glutamatergic neurons transmit signals from pontine REM-control regions to inhibitory glycinergic and GABAergic cells of the medulla that, in turn, suppress somatic motorneurons to produce REM atonia (Chase, 2008). Other peripheral signs triggered by activity in specific brainstem nuclei include penile and clitoral tumescence, fluctuations of autonomic functions such as heart rate and blood pressure, and skeletal muscle twitches.

The earliest cellular model of the brainstem REM sleep generator, the Reciprocal Interaction (RI) Model (Hobson, McCarley, & Wyzinski, 1975), suggested that regular alternation of REM and NREM sleep results from the interaction of aminergic REM-off and cholinergic REM-on neuronal populations in the pontine brainstem as illustrated in Figures 40.10A and B. This hypothesis was supported by extensive experimental and clinical findings showing, for example, that micro-injection of cholinergic agonists (or drugs like neostigmine that block the enzymatic degradation of naturally produced ACh) into the pontine ARAS of cats produced a state indistinguishable from naturally occurring REM sleep. In addition, REM sleep was reduced or blocked by micro-injection of noradrenergic or serotonergic agonists into similar sites. Moreover, in humans, REM is suppressed by systemic administration of drugs that block the synaptic re-uptake or enzymatic

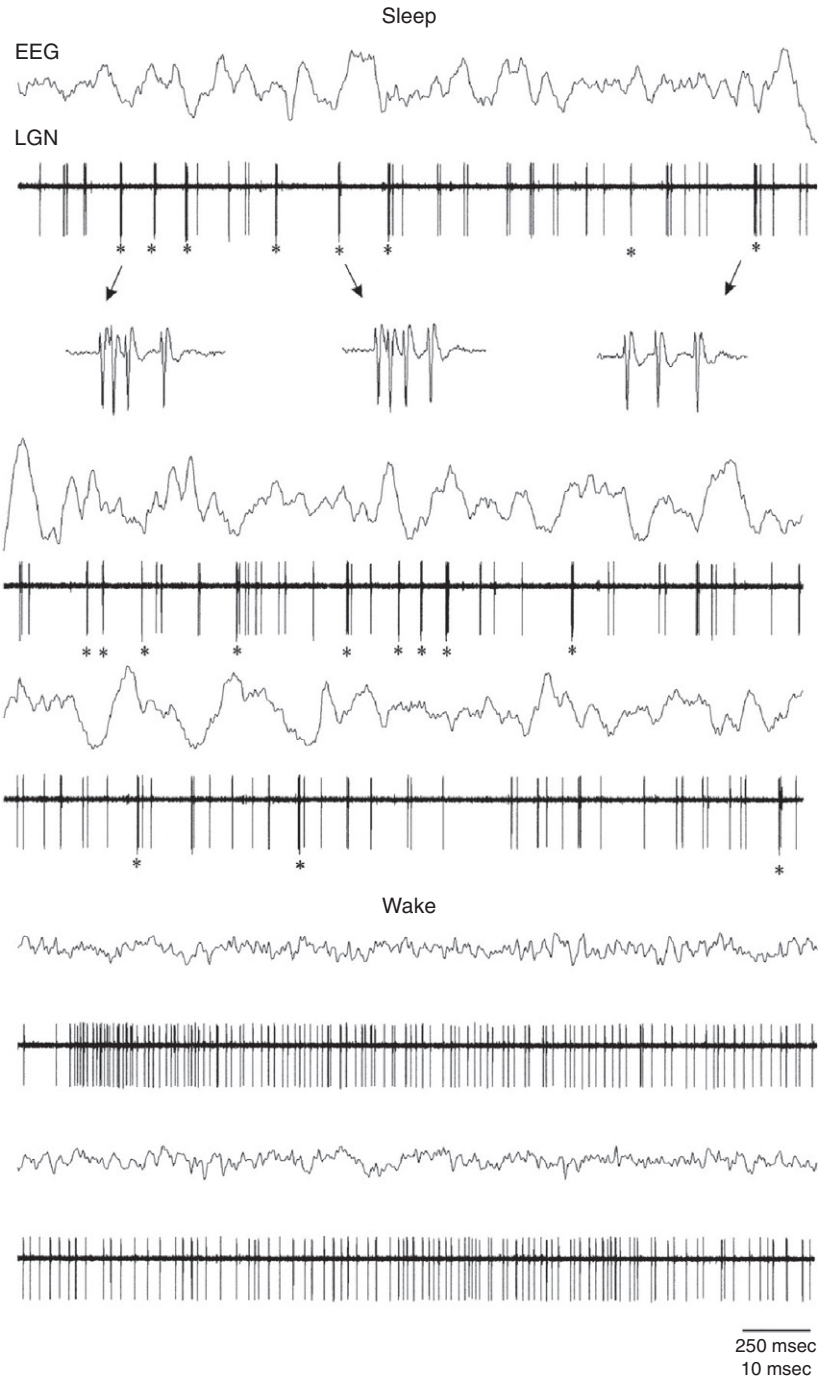


FIGURE 40.8 Two alternate firing modes of thalamic relay neurons: the bursting mode that is typical of slow wave sleep and the tonic (transmission) mode typical of waking. Traces show recordings in the thalamic lateral geniculate nucleus (LGN) of the cat along with simultaneous EEG from the occipital cortex. The top 7 traces show burst firing during sleep and the resultant synchronized low-frequency, high amplitude occipital EEG. The bottom 4 traces show recordings from these same sites during waking along with the resultant desynchronized, high-frequency EEG. Asterisks indicate bursting periods and the third trace shows an enlargement of the LGN channel during bursts. In the tonic mode during waking, firing frequency is proportional to the strength of a sensory input whereas during the bursting mode in sleep, transmission of sensory information to the cortex via the thalamus is blocked. For additional details on these modes of firing in thalamic relay neurons, as well as the membrane currents that underlie the rhythmic bursting mode, please see Chapter 5, Figure 5.14. *From Weyand et al. (2001),* J. Neurophysiology 85*:1107–18).*

degradation of naturally produced monoamines, such as many antidepressants.

Based upon subsequent reports, revision of the RI model added intermediate synaptic steps for both the activation of REM-on neurons and the suppression of REM-off aminergic nuclei (reviewed in Hobson et al., 2000). For example, mutually excitatory interactions between cholinergic and glutamatergic neurons may underlie the rapidly escalating

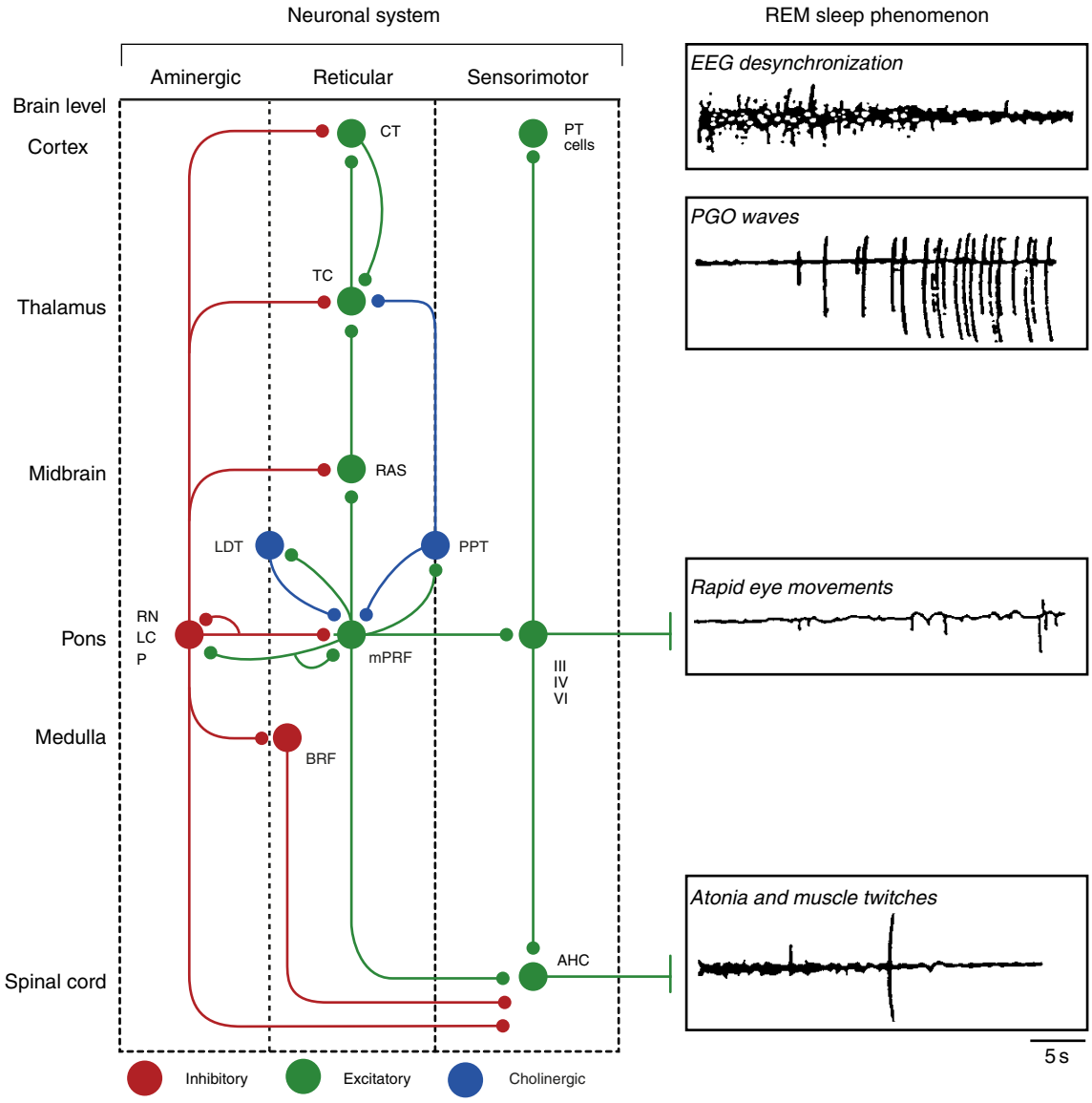


FIGURE 40.9 Schematic representation of the REM sleep generation process. A distributed network involves cells at brain levels from the spinal cord to the cortex (listed on the left). The network is represented as a diagram of three neuronal systems (aminergic, reticular, and sensorimotor) that mediate REM sleep phenomena (depicted in tracings on the right). Postulated inhibitory connections are shown as red circles; postulated excitatory connections as green circles; and cholinergic pontine nuclei are shown as blue circles. It should be noted that the actual synaptic signs of many of the aminergic and reticular pathways remain to be demonstrated, and, in many cases, the neuronal architecture is known to be far more complex than indicated here (e.g., the thalamus and cortex). During REM, additive facilitatory effects on pontine REM-on cells are postulated to occur via disinhibition (resulting from the marked reduction in firing rate by aminergic neurons at REM sleep onset) and through excitation (resulting from mutually excitatory cholinergic–noncholinergic cell interactions within the pontine tegmentum). The net result is strong tonic and phasic activation of reticular and sensorimotor neurons in REM sleep. REM sleep phenomena are postulated to be mediated as follows: EEG desynchronization results from a net tonic increase in reticular, thalamocortical, and cortical neuronal firing rates. PGO waves are the result of tonic disinhibition and phasic excitation of burst cells in the lateral pontomesencephalic tegmentum. Rapid eye movements are the consequence of phasic firing by reticular and vestibular cells; the latter (not shown) excite oculomotor neurons directly (see Chapter 32). Muscular atonia is the consequence of tonic postsynaptic inhibition of spinal anterior horn cells by the pontomedullary reticular formation. Muscle twitches occur when excitation by reticular and pyramidal tract motorneurons phasically overcomes the tonic inhibition of the anterior horn cells. RN, raphe nuclei; LC, locus coeruleus; P, peribrachial region; PPT, pedunculopontine tegmental nucleus; LDT, laterodorsal tegmental nucleus; mPRF, meso- and mediopontine tegmentum (e.g., gigantocellular tegmental field, parvocellular tegmental field); RAS, midbrain reticular activating system; BIRF, bulbospinal inhibitory reticular formation (e.g., gigantocellular tegmental field, parvocellular tegmental field, magnocellular tegmental field); TC, thalamocortical; CT, cortical; PT cell, pyramidal cell; III, oculomotor; IV, trochlear; V, trigmenial motor nuclei; AHC, anterior horn cell. *From Hobson et al. (2000).*

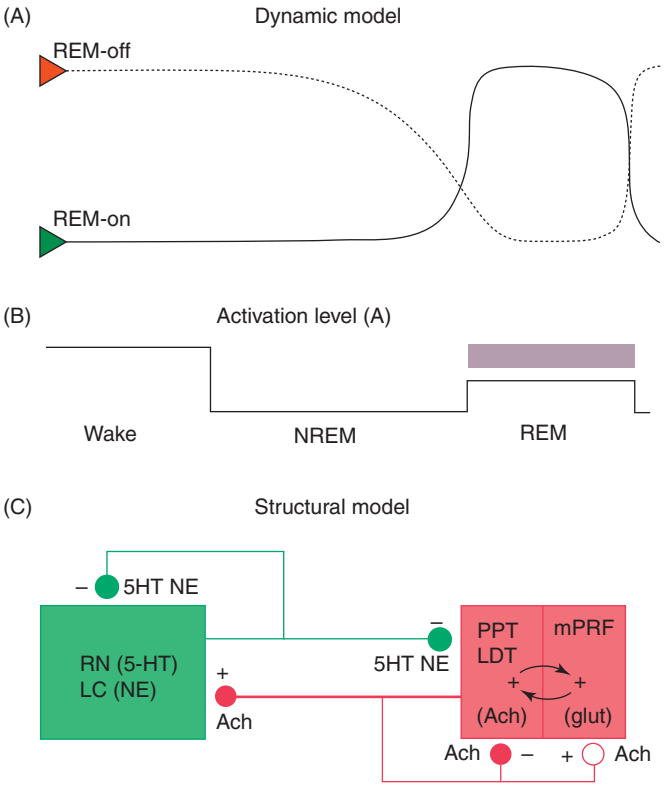


FIGURE 40.10 The Reciprocal Interaction (RI) Model (A) In the original RI model (Hobson et al., 1975; McCarley & Hobson, 1975), REM-on cholinergic neurons (Green triangle, solid line) both self-excite and excite aminergic REM-off neurons (Red triangle, dashed line). Aminergic REM-off neurons both inhibit cholinergic REM-on neurons and self-inhibit. This interaction leads to the alternation of behavioral states, depicted in (B), as follows: During waking, the REM-off aminergic system is active continuously and inhibits the pontine cholinergic system. During NREM sleep, aminergic activity decreases, allowing cholinergic activity to rise. The number of activated cholinergic REM-on neurons then exponentially increases due to self-excitation until sufficient numbers of aminergic REM-off neurons are also excited and begin to suppress the cholinergic REM-on population. However, these aminergic REM-off neurons then self-suppress, allowing the emergence of the next REM period. (C) In a subsequent revision of the RI hypothesis, mutually excitatory interactions between cholinergic and glutamatergic neurons underlie the rapidly escalating firing of pontine reticular REM-on neurons during REM sleep and cholinergic REM-on cells both self-excite and self-inhibit via cholinergic autoreceptors. *From Hobson et al. (2000).*

firing of pontine reticular REM-on neurons during REM sleep (Fig. 40.10C). Similarly, GABAergic inhibition by midbrain or local mesopontine interneurons may be responsible for inhibiting REM-off serotonergic and noradrenergic cells thereby allowing REM (Luppi et al., 2006). GABAergic inhibition of cholinergic cells in the PPT may prevent REM during SWS, while glutamatergic activation of a certain percentage of these cholinergic cells may trigger REM with further increase in percentage activated leading to waking (Datta & Maclean, 2007).

Other models on the control of REM onset and offset have emphasized interactions between brainstem GABAergic cell populations in which GABAergic inhibition plays both REM facilitatory and REM inhibitory roles. For example, Luppi et al. (2006) have suggested that the onset of REM in the rat is triggered by the release of noncholinergic, possibly glutamatergic, neurons of the pontine sublaterodorsal (SLD) nucleus from tonic GABAergic inhibition by midbrain reticular neurons that is present during NREM and wake. Most recently, Saper et al. (2010) have proposed that a GABAergic flip-flop switch, analogous to the above hypothalamic flip-flop switch, controls the alternation of REM and NREM sleep states in the rat. GABAergic REM-off regions in the ventrolateral periaqueductal gray and lateral pontine tegmentum project to and are mutually inhibitory with REM-on regions in the mesopontine SLD, precoeruleus, and periventricular gray areas. In this model, aminergic and cholinergic influences are modulatory, but a mutually inhibitory GABAergic flip-flop circuit is the final common pathway of REM-NREM switching. Aminergic inputs from the LC and DR excite the REM-off GABAergic regions and inhibit the REM-on ones, whereas cholinergic input from the LDT/PPT displays the opposite pattern. Similarly, orexinergic neurons excite the REM-off GABAergic regions, thereby inhibiting REM, whereas GABAergic projections from REM-promoting regions of the VLPO inhibit this same region. Discrepancies between models may result from species differences. For example, Datta and Maclean (2007) hypothesize different anatomical origins of pontine (PGO and P) waves in predator versus prey species.

Other Neurotransmitter Systems Participating in Behavioral State Control

Many additional neurotransmitter systems participate in the modulation of sleep states. Such modulatory neurochemicals may exert their effects on behavioral state by interacting with aminergic, cholinergic, GABAergic, and glutaminergic neurons that exert a more direct "executive" control over sleep and waking states, or they may mediate only specific physiological signs of REM and NREM sleep. Little attention was initially paid to dopamine (DA) in behavioral state control because, unlike the other monoamines, firing rates of dopaminergic cells in the SNpc and VTA did not appear to vary with behavioral state. However, synaptic levels of DA increase in REM when VTA neurons shift from tonic to burst firing modes (Dahan et al., 2007) and recently discovered wake-active dopaminergic neurons in the ventral PAG may constitute a dopaminergic component of ascending arousal analogous to serotonergic, noradrenergic, and histaminergic systems (Saper et al., 2010). In addition to the roles of galanin, orexin, and the cytokine peptides

described above, other neuropeptides implicated in the REM–NREM and sleep–wake cycles include vasoactive intestinal polypeptide, as well as numerous hormones, such as corticotropin releasing hormone (reviewed in Prospéro-García, Méndez-Díaz, Ruiz-Contreras, & Pérez-Morales, 2011). Nitric oxide (NO), an intercellularly diffusible gaseous neurotransmitter, is co-produced by cholinergic mesopontine neurons and may play a role in maintaining the cholinergically mediated REM sleep state in both the pons and the thalamus (Leonard & Lydic, 1999).

Molecular Biology of Behavioral State Control

Research on behavioral state control now extends inquiry beyond neurotransmitters and their receptors to intracellular second messenger cascades and gene transcription. The cyclic AMP-protein kinase A (cAMP-PKA) intracellular signaling pathway has been identified as the target of glutamatergic activation as well as GABAergic inhibition of PPT cholinergic cells generating physiological signs of REM (Datta & Maclean, 2007). For example, Bandyopadhya, Datta, and Saha (2006) have shown that intracellular concentrations of the catalytic subunit of PKA that enter the nucleus to phosphorylate CRE-binding protein vary with the prior amount of REM sleep. The specific profile of genes selectively activated in different sleep–wake states is also a current area of intense investigation (Cirelli, 2009), as is the genetic basis for variability in normal and disordered human sleep patterns (Landolt & Dijk, 2011). For example, genetic variation in adenosine deaminase, the degratory enzyme of adenosine, affects inter-individual variation in response to increased sleep pressure (Landolt & Dijk, 2011).

SYSTEMS NEUROSCIENCE OF SLEEP AND DREAMING

Electromagnetic Physiology

As described above, slow oscillatory rhythms produced in reentrant thalamo-cortical and cortico-cortico circuits predominate in human slow wave sleep (SWS). High-density EEG shows that the slow oscillation can originate in frontal regions of the brain and then travel posteriorly across the cortex (Massimini et al., 2004), as well as occur as a local cortical phenomenon (Nir et al., 2011). Frontal areas are also the first regions of the brain to show EEG slowing at sleep onset, as well as those that most reflect the increase in slow wave activity following sleep deprivation (Ferrara et al., 2002).

In contrast to NREM, wakelike electromagnetic brain activity in REM contains fast, gamma range (low gamma = 30–80 Hz, high gamma = 80–150 Hz) oscillations (Ferri et al., 2001). Gamma oscillations are associated with effortful waking cognitive activity such as working memory or attention to perceptual stimuli. Long-range synchronization of gamma oscillations across different brain regions has been proposed as one way in which the brain generates the coherent experience of a percept or memory across different sensory modalities ("binding"), and such binding may also create coherent hallucinatory percepts during REM-sleep dreams (Kahn, Pace-Schott, & Hobson, 1997). Interestingly, another difference in REM compared to waking, the loss of coherence (a quantitative measure of EEG synchrony) at gamma frequencies between frontal executive-control areas and posterior perceptual areas of the cortex, may be one mechanism by which dream bizarreness could arise (Corsi-Cabrera, Guevara, & Del Rio-Portilla, 2008).

Functional Neuroimaging of Sleep

Early positron emission tomography (PET) studies showed decreasing brain activation with sleep onset. As NREM deepens to SWS, global brain activity can decline by as much as 40% below waking (Maquet, 2000). Deactivation of the thalamus may occur at an early stage in this process and contribute to subsequent deactivation of widespread cortical, limbic, and striatal areas (Dang-Vu et al., 2010). fMRI has further demonstrated decreases in perfusion initially of frontal cortical areas and, as NREM deepens, also in the hypothalamus, striatum, thalamus, hippocampal formation, and posterior cortex (Kaufmann et al., 2006).

Recent studies have examined the effect of NREM sleep on activity in the brain's default network, a group of (mostly) midline brain areas that characteristically show a *decrease* in regional cerebral blood flow signal in response to a wide variety of goal-directed tasks, especially those requiring attention to exteroceptive (i.e., outside the body), perceptual stimuli (Buckner, Andrews-Hanna, & Schacter, 2008). In the absence of exteroceptive attention, these regions are believed to carry out adaptive, homeostatic, interoceptive (directed toward the internal milieu), and self-related functions, including imaginal simulation of future needs and desires, retrieval of autobiographical episodic memory, and self-regulation. In the blood oxygen level dependent (BOLD) signal of fMRI, the default network manifests as temporal synchrony of low-frequency (0.01–0.1 Hz) oscillations among its different regions (Buckner et al., 2008). This signature synchrony of BOLD fluctuations that defines the default mode during task-free waking persists into early stages of NREM sleep (N1, N2) but is lost during SWS (Samann et al., 2011).

With transition from NREM to REM, most lateral portions of the cortex remain less active than in waking, whereas activation, sometimes to levels exceeding that of

waking, occurs in the pons, midbrain, thalamus, amygdala, hypothalamus, and basal ganglia, as well as limbic-related cortices (reviewed in Pace-Schott, 2011b). These regions comprise what Nofzinger et al. (2004) call the "anterior paralimbic REM activation area," and these findings have led to an important expansion of theories on the neural origin of REM dreams, as described in the next section.

A Neurobiological Model of Dreaming

From evidence such as the above, we can suggest a neurobiological model of REM-sleep dreaming—the major conscious state that humans experience during sleep. Although REM dreams—being the most vivid, motoric, emotional, and bizarre—are modeled, dreaming also occurs at lower intensity and rate of recall during NREM (Hobson et al., 2000). (More detailed presentations of this model, shown in Fig. 40.11, can be found in Pace-Schott, 2011b.) Table 40.1 lists the phenomenological features of dreams that this neurobiological model seeks to explain.

In REM, the forebrain is activated, as in waking, by ARAS cholinergic nuclei and their brainstem targets that, by activating the thalamus and basal forebrain, allow these forebrain structures to, in turn, activate the limbic system and cortex via glutamatergic and cholinergic mechanisms, respectively. As originally hypothesized in the Activation-Synthesis Hypothesis (Hobson & McCarley, 1977), such forebrain activation results in a markedly different sort of conscious experience owing to the almost complete lack

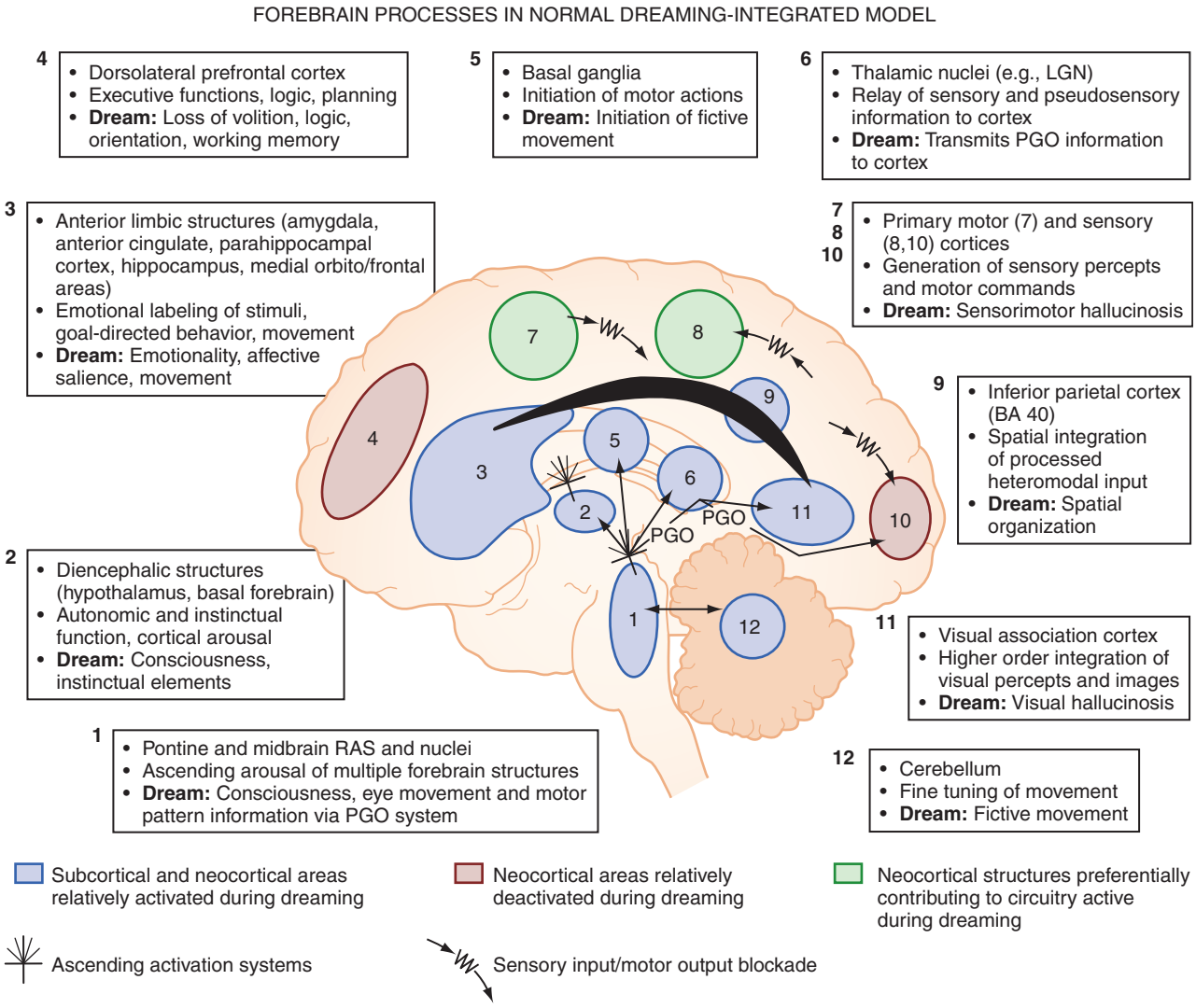

FOREBRAIN PROCESSES IN NORMAL DREAMING-INTEGRATED MODEL
4
• Dorsolateral prefrontal cortex
• Executive functions, logic, planning
• Dream: Loss of volition, logic, orientation, working memory
5
• Basal ganglia
• Initiation of motor actions
• Dream: Initiation of fictive movement
6
• Thalamic nuclei (e.g., LGN)
• Relay of sensory and pseudosensory information to cortex
• Dream: Transmits PGO information to cortex
3
• Anterior limbic structures (amygdala, anterior cingulate, parahippocampal cortex, hippocampus, medial orbito/frontal areas)
• Emotional labeling of stimuli, goal-directed behavior, movement
• Dream: Emotionality, affective salience, movement
7 8 10
• Primary motor (7) and sensory (8,10) cortices
• Generation of sensory percepts and motor commands
• Dream: Sensorimotor hallucinosis
9
• Inferior parietal cortex (BA 40)
• Spatial integration of processed heteromodal input
• Dream: Spatial organization
2
• Diencephalic structures (hypothalamus, basal forebrain)
• Autonomic and instinctual function, cortical arousal
• Dream: Consciousness, instinctual elements
11
• Visual association cortex
• Higher order integration of visual percepts and images
• Dream: Visual hallucinosis
1
• Pontine and midbrain RAS and nuclei
• Ascending arousal of multiple forebrain structures
• Dream: Consciousness, eye movement and motor pattern information via PGO system
12
• Cerebellum
• Fine tuning of movement
• Dream: Fictive movement
PGO
PGO
Subcortical and neocortical areas relatively activated during dreaming
Neocortical areas relatively deactivated during dreaming
Neocortical structures preferentially contributing to circuitry active during dreaming
Ascending activation systems
Sensory input/motor output blockade

FIGURE 40.11 Forebrain processes in normal dreaming—an integration of neurophysiological, neuropsychological, and neuroimaging data. Regions 1 and 2: ascending arousal systems; 3: subcortical and cortical limbic and paralimbic structures; 4: dorsolateral prefrontal executive association cortex; 5: motor initiation and control centers; 6: thalamocortical relay centers and thalamic subcortical circuitry; 7: primary motor cortex; 8: primary sensory cortex; 9: inferior parietal lobe; 10: primary visual cortex; 11: visual association cortex; 12: cerebellum.

TABLE 40.1 Phylogeny of Rest and Sleep

Organism	Rest	Sleep	REM Sleep
Mammals			
Adults	+	+	+
Neonates	+	+	+
Birds			
Adults	+	+	+
Neonates	+	+	+, –
Reptiles	+	+	–
Amphibians	+	–	–
Fish	+	+, –	–

Note: REM, rapid eye movement; +, present; +, –, ambiguously or inconsistently present; –, absent.

of noradrenergic, serotonergic, and histaminergic neuromodulation of the cortex. Notably, as detailed in Chapter 46 and 50, aminergic modulation facilitates higher cognitive functions such as attention and working memory.

Activity in the "anterior paralimbic REM activation area" may generate the emotionally salient features of dreaming. For example, activity in the amygdala, related basal forebrain areas (e.g., bed nucleus of the stria terminalis) and medial prefrontal (e.g., subgenual) areas, which are components of fear-processing circuits, may result in anxious or fearful dreams (Levin & Nielsen, 2007). On the other hand, activity in the midbrain, ventral striatum, and pallidum (components of reward circuits) may produce appetitive dreams (Solms, 1997). Similarly, activity in the hypothalamus and brainstem (e.g., periaqueductal gray) may trigger instinctual programs resulting in sexual, panic, or rage experiences (Jouvet, 1999). In addition, the richly social and interpersonally salient nature of dreams may result from activation of medial prefrontal areas involved in social cognition, such as the capacity for forming a "theory of mind," empathizing, or moralizing.

Veridical replay of actual episodic experiences is quite rare in dreams (Stickgold, Hobson, Fosse, & Fosse, 2001), but recognition memory can sometimes be bizarrely enhanced, such as when we "just know" who a dream character is despite marked differences from the real-life counterpart (Kahn, Stickgold, Pace-Schott, & Hobson, 2000). Such credulity in dreams may be one manifestation of how low activity in lateral prefrontal areas results in dream bizarreness. Ventromedial frontal areas active in REM are involved in rapid, feelings-based appraisal of memory accuracy and may accept dream events without the verification provided by lateral and polar prefrontal areas that are needed for critical thinking and strategic search of memory. In a similar manner, other forms of dream bizarreness, such as incongruous juxtapositions, sudden discontinuities of plot, or bizarre uncertainty (see Hobson, 1988), may be uncritically accepted due to deactivation of lateral and polar prefrontal cortex. Interestingly, damage to lateral frontal areas has little effect on dreaming (Solms, 1997).

Although dreams abound with hallucinatory motion, REM atonia prevents contraction of skeletal muscles for the enactment of motor commands except in the case of REM behavior disorder (see below). However, REM-related activity in basal ganglia may produce the sense of initiated motion, while activity in the cerebellum may produce vestibular sensations such as flying, falling, or moving.

The Activation-Synthesis Hypothesis suggested that the striking visual hallucinations of dreams arise when phasic signals from the brainstem during REM, such as PGO waves, arrive at the cortex through a dedicated visual pathway and are interpreted by the visual cortex as true sensory input (Hobson & McCarley, 1977). Activation of visual association areas during REM by such ascending signals may allow emergence of complex hallucinations of actual objects and persons (Hobson et al., 2000). For example, regions of the ventral stream of visual processing (see Chapter 44), such as the fusiform face area or the parahippocampal place area, increase their activity during REM (Braun et al., 1998). Interestingly, damage to such ventral stream areas can result in nonvisual dreams (Solms, 1997).

Much, however, remains unexplained. For example, why are dream hallucinations integrated into elaborate plots in the absence of higher-order functions that, in waking, would be needed to create a coherent story? Similarly, why should damage to the inferior parietal cortex result in dream cessation when this area is among those that remain deactivated during REM? And, of course, how are dreams generated in NREM sleep?

DEVELOPMENT AND PHYLOGENY OF SLEEP

Changes in Sleep over the Human Lifespan

Human sleep shows striking developmental changes across the lifespan (Fig. 40.12). Newborns spend 17 to 18 hours per day sleeping in 3- to 4-hour bouts that, by approximately age 6, have consolidated into a single 10- to 12-hour nocturnal sleep period (Scher, 2008). An important transition during this period involves consolidation of diurnal sleep, first into a single daytime nap, that is then abandoned for solely nocturnal sleep (Davis, Parker, & Montgomery, 2004). SWS is absent at birth and develops during the first year of life, as do sleep spindles. In contrast, REM makes up around 50% of sleep in newborns, who also display incomplete atonia with a poorly differentiated EEG. By age 4, however, REM has declined to approximately adult levels of 20 to 25%. As the preadolescent enters adolescence, there is an endogenous circadian

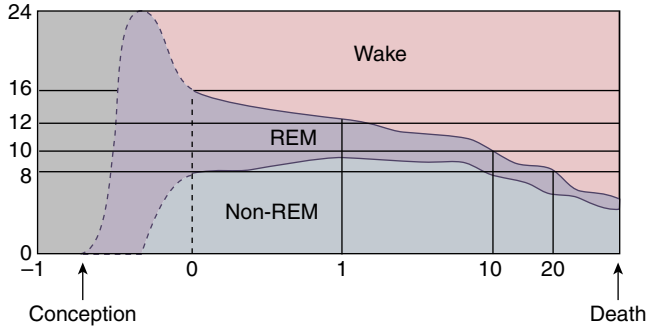


FIGURE 40.12 Portions of a 24-hour day that are devoted to waking, REM sleep, and non-REM (NREM) sleep change over a lifetime. Although the timing of these changes *in utero* is not known with certainty (dotted lines), data from premature infants are consistent with REM sleep occupying most of life at a gestational age of 26 weeks. After 26 weeks, the time spent in waking increases until death.

phase delay resulting in later sleep onset and rising time (Carskadon, 2011) as well as a decline in SWS. In young adults, on average, total sleep time is around 6 to 8 hours, of which about 5% is N1, 50 to 60% N2, 15 to 20% SWS, and 20 to 25% REM (Ohayon, Carskadon, Guilleminault, & Vitiello, 2004). Healthy aging is characterized by a diminished quality and duration of sleep, with increased time awake after sleep onset (Pace-Schott & Spencer, 2011). Older adults awaken more frequently and show a dramatic decline in SWS along with less marked decreases in numbers of sleep spindles and K-complexes. Opposite to changes entering adolescence, circadian rhythms of temperature, melatonin, and cortisol become phase advanced (earlier sleep and rising) in older adults.

Phylogeny of Sleep

All mammals display REM and NREM sleep, but the duration and architecture of sleep vary widely. In general terms, smaller mammal species sleep more than larger ones, brain weight positively correlates with length of the REM-NREM cycle, and those born less mature have more REM (Hobson, 1989). In some marine mammals and birds, sleep occurs in one hemisphere at a time which, in the case of cetaceans, allows continued breathing (Rattenborg, Amlaner, & Lima, 2000). Birds have very short sleep and REM/NREM cycles, with minimal REM lacking atonia. Reptiles, amphibia, fish, and invertebrates have periods of quiescence with both circadian and seasonal periodicities, but the equivalence of these states to the sleep of homeotherms is just now being elucidated (Allada & Siegel, 2008). As noted in Chapter 39, the genetic basis of circadian rhythms is remarkably conserved from the most primitive to the most advanced phyla. Table 40.2 shows what is currently believed true about sleep states in the various phyla.

SLEEP DISORDERS

Disrupted Control of Behavioral State Can Lead to Insomnia and Hypersomnia

When brain aminergic activity is excessive due to stress, disease, or genetic predisposition, behavioral state is shifted in the direction of hyperarousal, and insomnia can result. Behavioral and pharmacologic interventions can be used to reduce this hyperarousal (Ellenbogen & Pace-Schott, 2012). Among pharmacological treatments, the most common are drugs such as zolpidem that target the benzodiazepine binding site on $GABA_A$ receptors. Binding to the benzodiazepine receptor site increases the ability of GABA to trigger chloride influx through the ionotropic $GABA_A$ receptor. The resulting hyperpolarization of the membrane then leads to neuronal inhibition and behavioral sedation. As described in Chapter 8, the $GABA_A$ receptor is a pentameric structure whose subunits exist in various isoforms. By targeting specific isoforms, hypnotic drugs have been developed that promote sleep with a lesser tendency to be addictive or to cause cognitive problems compared to "traditional benzodiazepines" like valium.

Abnormal decreases in CNS aminergic activity have the opposite effect on cognitive and behavioral systems. In the case of narcolepsy, the reduced orexinergic drive on aminergic systems allows sleep attacks. In such conditions, the REM generator is disinhibited, as can be seen in REM sleep attacks of narcoleptics. These symptoms can be decreased by aminergic agonists, such as antidepressants and psychostimulants, that block amine reuptake. The drug now most commonly used to treat hypersomnolence, modafinil (and its R-enantiomer armodafinil), shows weak reuptake blockade of norepinephrine and dopamine but may also augment electrical coupling among cells of the ARAS (Ellenbogen & Pace-Schott, 2012).

Respiratory Dysfunction Is Common in Sleep

One of the more dramatic problems associated with sleep is an exaggeration of the normal decline in drive from brainstem respiratory centers that occurs during sleep, a condition known as central sleep apnea (Fig. 40.13). A much more common condition, obstructive or peripheral sleep apnea (OSA), results from mechanical collapse or compression of the airway, a condition especially common in obese people (see Kryger et al., 2011, for extensive reviews; Fig. 40.13). In OSA, central respiratory drive causes continued respiratory effort. Eventually, the increase of CO_2 in the brain leads to an arousal that in turn produces increased muscle tone in the oropharynx, and the airway opens. As the person

TABLE 40.2 Physiological Basis of Changes That Occur during Dreaming

Function	Change (Compared with Waking)	Hypothesized Cause
Sensory input	Blocked	Presynaptic inhibition
Perception (external)	Diminished	Blockade of sensory input
Perception (internal)	Enhanced	Removal of inhibition from networks that store sensory representations
Attention	Lost	Aminergic modulation decreases, causing a decrease in the ratio of signal to noise
Memory (recent)	Diminished	Because of a decrease in aminergic activity, activated representations are not stored in memory
Memory (remote)	Enhanced	Removal of inhibition from networks that store memory representations
Orientation	Unstable	Internally inconsistent signals are generated by cholinergic systems
Thought	Poor reasoning, processing hyper-associative	Loss of attention, memory, and volition leads to failure of sequencing and rule inconstancy; analogy replaces analysis
Insight	Self-reflection lost	Failures of attention, logic, and memory weaken second- (and third-) order representations
Language (internal)	Confabulatory	Aminergic demodulation frees the use of language from the restraint of logic
Emotion	Episodically strong	Cholinergic hyperstimulation of the amygdala and related structures of the temporal lobe triggers emotional storms, which are not modulated by aminergic activity
Instinct	Episodically strong	Cholinergic hyperstimulation of the hypothalamus and limbic forebrain triggers fixed motor programs, which are experienced fictively but not enacted
Volition	Weak	Cortical motor control cannot compete with disinhibited subcortical networks
Output	Blocked	Postsynaptic inhibition

falls into deeper sleep again, the oropharynx relaxes and the airway can become obstructed again until arousal recurs. In severe cases, arousals every one to two minutes all night are common.

People with sleep apnea usually feel excessively sleepy during the daytime because frequent arousal prevents deep, sustained sleep. They are often unaware of their sleep-dependent breathing pattern because they are never fully aroused from sleep. Instead, observant friends or family often describe seeing these signs of respiratory dysfunction. In a sleep laboratory, EEG, respiratory effort, and cardiovascular parameters (e.g., heart rate, blood pressure, and oxygen saturation) are measured to quantify the problem. OSA can be treated by having the person wear a continuous positive airway (CPAP) mask through which air is forced gently, helping to keep the airways open. The pressure of the air fluctuates to allow regular exhalation. People with OSA are treated not only to alleviate their constant drowsiness but to prevent long-term effects of sleep apnea, such as cardiopulmonary complications and cognitive impairment.

Motor Disturbances Can Occur during Sleep

Parasomnias that affect the skeletal motor system can occur in either REM or NREM sleep. In NREM sleep, these parasomnias (e.g., sleep walking, sleep talking, tooth grinding, and night terrors) are mainly seen in young people, and reflect motor responses to the outputs of central motor pattern generators. Normally, responses to these generators are disinhibited at sleep onset and inhibited during REM sleep (by the REM atonia previously described). Excessive motor output during REM, termed REM sleep behavior disorder (RBD), can be an early sign of degenerative brain disease in the elderly when the brainstem neural circuits controlling REM atonia have become damaged. Because of a failure to inhibit the expression of motor acts in REM, people with this disorder may literally enact their dream scenarios in their bedrooms. A syndrome strikingly similar to RBD can be induced experimentally by bilateral lesions of the pontine tegmentum in cats. When in REM sleep, these animals perform stereotyped behavior sequences, such as hissing, piloerection, pouncing, and jumping. Jouvet called these

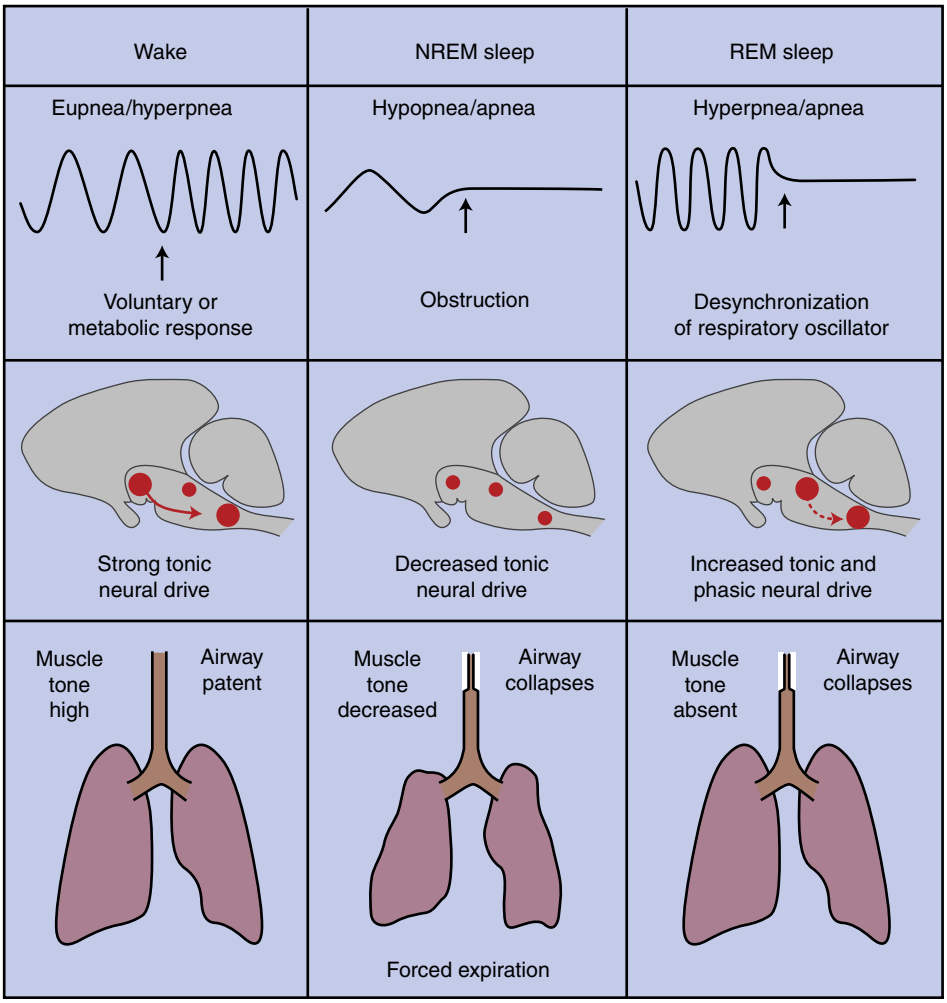


FIGURE 40.13 Central and obstructive sleep apnea. During waking, the respiratory oscillator of the medulla receives tonic drive from other neural structures and can respond to voluntary and metabolic signals to change breathing pattern. Muscle tone keeps the oropharynx open to the flow of air. In NREM sleep, central drive decreases, and the rate and depth of ventilation fall. If the decrease is excessive, central sleep apnea results, including a complete, albeit temporary, cessation of breathing. If the airway collapses, obstructive sleep apnea may result, with a similar cessation of breathing. During REM sleep, activation of pontine generator neurons drives the respiratory oscillator, and desynchronization may lead to breathing efforts that are too frequent or strong (hyperpnea) or that stop. During REM sleep the oscillator also becomes unresponsive to metabolic signals.

"hallucinatory behaviors," and Morrison called the phenomenon "REM sleep without atonia" (see Jouvet, 1999). Because the ability of a cat to inhibit motor pattern commands during REM sleep is impaired by the lesions, they act out commands that are normally initiated during REM sleep. The instinctual aspect of these released behaviors is relevant to the psychophysiology of dreaming as described above.

Other Sensorimotor Disturbances

Among the most interesting of sleep disorders is restless leg syndrome (RLS), a disorder that is often accompanied by another sleep disorder, periodic limb movement disorder (PLM). In RLS, while trying to go to sleep, the patient experiences a highly unpleasant burning or crawling sensation in the legs that can be relieved only by movement, often by arising and pacing. RLS is treated with dopaminergic agents that compensate for deficient output from dopaminergic neurons of the A11 region of the posterior hypothalamus that is the sole source of spinal cord dopamine (Paulus & Schomburg, 2006). Dopamine regulates sensory input to the dorsal horn of the spinal cord as well as the transmission of this input to the brain. In RLS, dopamine deficiency may allow excessive transmission of the normal low-level activity entering the spinal cord from peripheral receptors that, when transmitted to the brain, then produces abnormal sensations. Similarly, dopaminergic medications treat PLM by boosting dopaminergic inhibition of spinal early flexor reflexes that, when such inhibition is deficient, can be abnormally triggered by such baseline sensory inputs.

Disorders of Circadian Misalignment

Misalignment of waking activities with endogenous circadian rhythms is a common source of disordered sleep. As has been discussed, abnormal delay of sleep onset and offset (delayed sleep phase syndrome) is more common in adolescents and young adults, whereas abnormally early sleep onset and offset (advanced sleep phase syndrome) is more common in the elderly. In addition, overnight shift work or travel across time zones can lead to a desired sleep bout becoming misaligned with one's

endogenous circadian propensity for sleep and waking. Because light can entrain the circadian oscillator in the SCN, precisely timed treatment with bright light is the most commonly employed strategy for treating those circadian disorders. Melatonin and its agonists are also used to promote sleep onset in disorders of circadian misalignment.

THE PURPOSE OF SLEEP

Sleep Appears to Have Multiple Functions

As a quiescent and ecologically protected behavior, sleep fosters conservation of energy, defense from predation, and an opportunity for repair of injury. The anabolic character of NREM sleep (e.g., brain and body inactivity and hormone release) suggests a rest and restoration function. For example, Benington and Heller (1995) propose that an important function of NREM sleep is to restore brain glycogen stores. In contrast, the high proportion of REM sleep in the developing brain (Fig. 40.12) suggests a role of REM sleep in brain development and plasticity (Marks, Shaffery, Oksenberg, Speciale, & Roffwarg, 1995).

The crucial anabolic and conservative role of sleep is illustrated by studies showing that prolonged sleep deprivation results in the disruption of metabolic and caloric homeostasis and eventually death (Rechtschaffen & Bergmann, 2002). In rats, sleep deprivation was fatal when it persisted for 4 to 6 weeks. Early in the deprivation period, the rats began to eat more but could not maintain their body weight. Later, they lost their ability to maintain body temperature, developed strong heat-seeking behavior, and finally died. It is suspected their death resulted from overwhelming sepsis because of immunodeficiency. The linkage between sleep and immune function is further emphasized by their numerous interactions (Imeri & Opp, 2009). For example, NREM sleep is enhanced by the cytokines interleukin-1 and interleukin-2, both of which are released during NREM sleep (Krueger, Rector, & Churchill, 2007).

Sleep also plays a key role in hormonal regulation. Growth and development are enhanced by the release of growth hormone and gonadotropins during SWS (Leproult & Van Cauter, 2010). Such hormone release declines, along with SWS, after growth and sexual maturation are complete (Van Cauter, Leproult, & Plat, 2000). In addition, sleep loss has been linked to dysregulation of glucose metabolism and increased risk of the metabolic syndrome diabetes and obesity (Van Cauter et al., 2007).

Several recent theories extend the role of sleep to the cellular and molecular levels. The synaptic homeostasis theory (Tononi & Cirelli, 2006) suggests that sleep, especially SWS, may serve to reduce or "downscale" excessive synaptic potentiation produced during prior waking, eliminating the more weakly potentiated synaptic connections and leaving only those associated with the most salient waking events. A particularly intriguing idea is that sleep is necessary to optimize correct folding of proteins in the endoplasmic reticulum (ER), as evidenced by the fact that sleep deprivation increases production of chaperone proteins involved in the ER stress (or "unfolded protein") response (Naidoo, 2009). As further detailed in the next section, sleep and memory consolidation are closely interconnected, and the cAMP-PKA-CREB, second messenger cascade, important to memory formation in wake, also appears to be important in sleep-related memory consolidation (Hernandez & Abel, 2011).

Sleep and Memory Are Interdependent

Perhaps the clearest evidence of a function for sleep is its role in consolidating recently formed memories (for recent reviews, see Diekelmann & Born, 2010; Walker & Stickgold, 2006). One of the most notable examples is the discovery, in rats, by Carlyle Smith, of "REM windows"

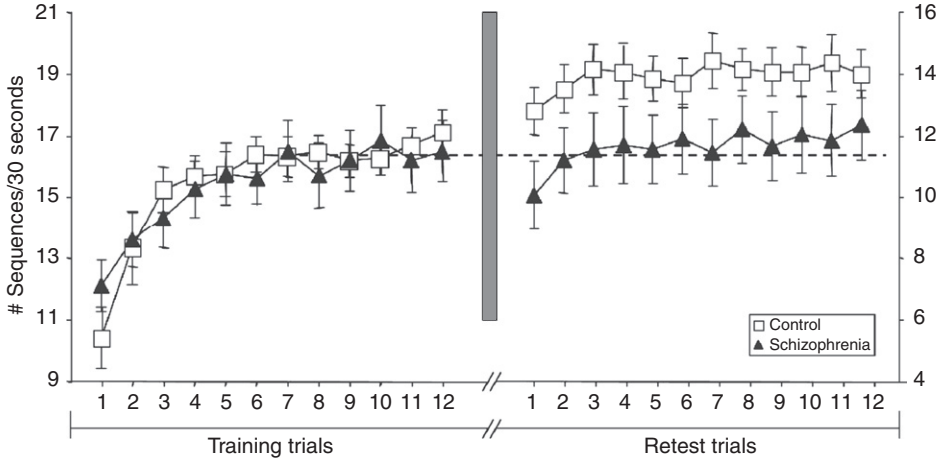


FIGURE 40.14 Sleep-dependent memory consolidation. Subjects were asked to type the sequence "41324" as quickly and accurately as they could for 30 seconds and then rest for 30 seconds before doing it again. Twelve trials were performed on each of two successive days. While the control subjects showed a 16% increase in speed overnight, the schizophrenia patients showed none.

(see Smith, 1995), which are time periods after an animal is trained on specific tasks when the animal shows enhanced quantities of REM sleep and when retention of learning could be decreased by selective REM sleep deprivation. Thus, not only does REM sleep appear to be critical for the maintenance of this learning, but animals also increase their REM sleep after training, presumably by a homeostatic mechanism that regulates the amounts of different sleep stages based on the needs of the organism. Subsequent studies in humans have led to the theory that REM sleep is especially important for the consolidation of procedural memories, which are defined as memories of how to do things in the absence of explicit rules, such as how to ride a bicycle or read a map (Walker & Stickgold, 2006). In contrast, SWS has been implicated in the consolidation of explicit or declarative memory, which includes autobiographical memories as well as general knowledge of facts (Diekelmann & Born, 2010). However, studies on a procedural visual discrimination task have shown a requirement for both REM and SWS on the night following training in order for next-day improvement on the task to be observed (Stickgold, Whidbee, Schirmer, Patel, & Hobson, 2000b). This finding supports an earlier theory based on animal studies in which a two-step process involving both REM and SWS was hypothesized. In humans, the requirement for sleep immediately following learning was shown when subjects who were sleep deprived the night following training, but who then were allowed two nights of unrestricted recovery sleep before testing, showed no significant task improvement (Stickgold, James, & Hobson, 2000a). In addition to procedural learning tasks, REM effects have been reported in humans on learning of complex logic games, foreign language acquisition, intensive studying in general, and both REM and SWS effects have been reported for the consolidation and processing of emotional memories. In contrast, sleep-dependent consolidation of simple motor skill learning, such as for a sequential finger tapping task, correlates with Stage II NREM sleep (Walker, Brakefield, Morgan, Hobson, & Stickgold, 2002).

The Functions of Dreaming, If Any, Are Poorly Understood

Does dreaming also serve a function? Answering this question is not yet possible, because we do not know how to separate possible functions of the experience of dreaming from functions of the neurobiological processes that produce it. But REM sleep dream content can predict who will become clinically depressed following marital divorce or separation (Cartwright, Young, Mercer, & Bears, 1998), suggesting that dreaming may serve to modulate emotional states during waking.

Dreaming may also play a role in modifying associative memory networks. Subjects awakened from REM sleep show preferential activation of more distantly related associated than seen either during waking or following NREM sleep (Stickgold et al., 1999). Such preferential activation of weak associations might explain the bizarre content of dreams during REM sleep and may permit the brain to discover creative solutions to problems during sleep. Indeed, sleep has been shown to increase the likelihood of such discoveries (Fig. 40.15; Wagner, Gais, Haider, Verleger, & Born, 2004). Most recently, it has been demonstrated that individuals who dream about a cognitive task following practice perform better on that task when they awaken (Wamsley, Tucker, Payne, Benavides, & Stickgold, 2010). But whether the conscious experience of dreaming is crucial for any of these sleep-dependent processes remains unknown.

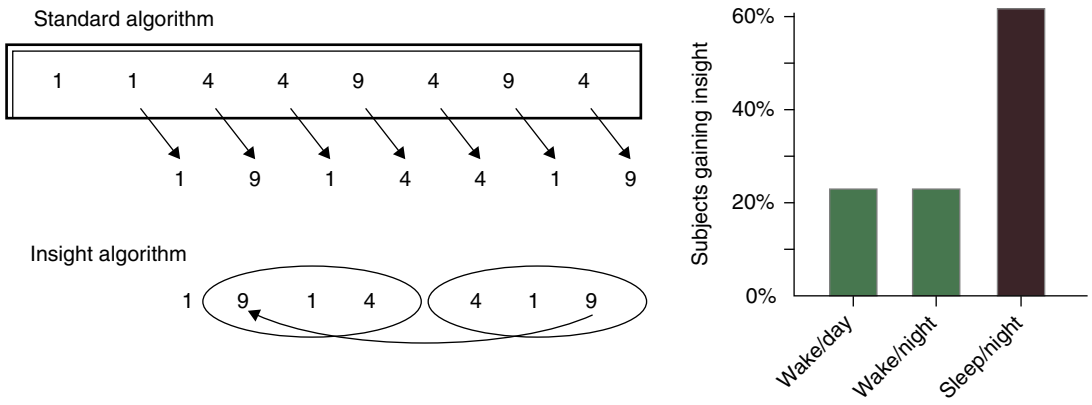


FIGURE 40.15 Sleep facilitates mathematical insight. Subjects are taught to reduce an 8-digit sequence to a single digit (9 at the right), through six intermediate calculations (in italics) with a standard algorithm. Unknown to the subjects, the task is designed so that the last three calculations form a mirror image of the preceding three, and thus the second intermediate calculation matches the final answer. Right: Subjects who slept between training and testing were more than twice as likely to discover the insight algorithm than those not allowed to sleep.

Dreaming may be viewed as evidence for a state of primary consciousness whose function is to provide the brain with an updatable virtual reality model of the brain that it uses in waking consciousness (Hobson, 2009). This view is surprisingly compatible with thermodynamic hypotheses about the brain (Friston, 2010).

References

Alam, M. N., Szymusiak, R., & Mcginty, D. (2011). Hypocretinergic system: Role in REM-sleep regulation. In B. N. Mallick, S. R. Pandi-Permual, R. W. Mccarley, & A. R. Morrison (Eds.), *Rapid eye movement sleep, regulation and function*. Cambridge: Cambridge University Press.

Allada, R., & Siegel, J. M. (2008). Unearthing the phylogenetic roots of sleep. *Current Biology, 18,* R670–R679.

Amzica, F., Steriade, M. (1996). Progressive cortical synchronization of ponto-geniculo-occipital potentials during rapid eye movement sleep. *Neuroscience, 72,* 309–314.

Arendt, J. (2006). Melatonin and human rhythms. *Chronobiology International, 23,* 21–37.

Aserinsky, E., & Kleitman, N. (1953). Regularly occurring periods of eye motility, and concomitant phenomena, during sleep. *Science, 118,* 273–274.

Bandyopadhya, R. S., Datta, S., & Saha, S. (2006). Activation of pedunculopontine tegmental protein kinase A: A mechanism for rapid eye movement sleep generation in the freely moving rat. *Journal of Neuroscience, 26*(35), 8931–8942.

Benington, J. H., & Heller, H. C. (1995). Restoration of brain energy metabolism as the function of sleep. *Progress in Neurobiology, 45,* 347–360.

Bersagliere, A., & Achermann, P. (2010). Slow oscillations in human non-rapid eye movement sleep electroencephalogram: effects of increased sleep pressure. *Journal of Sleep Research, 19,* 228–237.

Bes, F., Jobert, M., & Schulz, H. (2009). Modeling napping, post-lunch dip, and other variations in human sleep propensity. *Sleep, 32,* 392–398.

Borbely, A. A. (1982). A two process model of sleep regulation. *Human Neurobiology, 1,* 195–204.

Braun, A. R., Balkin, T. J., Wesensten, N. J., Gwadry, F., Carson, R. E., Varga, M., … Herscovitch, P. (1998). Dissociated pattern of activity in visual cortices and their projections during human rapid eye movement sleep. *Science, 279,* 91–95.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Annals of the New York Academy of Science, 1124,* 1–38.

Carskadon, M. A. (2011). Sleep in adolescents: The perfect storm. *Pediatric Clinics of North America, 58,* 637–647.

Cartwright, R., Young, M. A., Mercer, P., & Bears, M. (1998). Role of REM sleep and dream variables in the prediction of remission from depression. *Psychiatry Research, 80,* 249–255.

Chase, M. H. (2008). Confirmation of the consensus that glycinergic postsynaptic inhibition is responsible for the atonia of REM sleep. *Sleep, 31,* 1487–1491.

Cirelli, C. (2009). The genetic and molecular regulation of sleep: From fruit flies to humans. *Nature Reviews, 10,* 549–560.

Colrain, I. M. (2005). The K-complex: A 7-decade history. *Sleep, 28,* 255–273.

Corsi-Cabrera, M., Guevara, M. A., & Del Rio-Portilla, Y. (2008). Brain activity and temporal coupling related to eye movements during REM sleep: EEG and MEG results. *Brain Research, 1235,* 82–91.

Crunelli, V., & Hughes, S. W. (2010). The slow (<1 Hz) rhythm of non-REM sleep: A dialogue between three cardinal oscillators. *Nature Neuroscience, 13,* 9–17.

Czeisler, C. A., Duffy, J. F., Shanahan, T. L., Brown, E. N., Mitchell, J. F., Rimmer, D. W., … Kronauer, R. E. (1999). Stability, precision, and near-24-hour period of the human circadian pacemaker. *Science, 284,* 2177–2181.

Daan, S., Beersma, D. G. M., & Borbely, A. A. (1984). Timing of human sleep: Recovery process gated by a circadian pacemaker. *American Journal of Physiology, 246,* R161–R178.

Dahan, L., Astier, B., Vautrelle, N., Urbain, N., Kocsis, B., & Chouvet, G. (2007). Prominent burst firing of dopaminergic neurons in the ventral tegmental area during paradoxical sleep. *Neuropsychopharmacology, 32,* 1232–1241.

Dahlstrom, A., & Fuxe, K. (1964). Localization of monoamines in the lower brain stem. *Experientia, 20,* 398–399.

Dang-Vu, T. T., Schabus, M., Desseilles, M., Sterpenich, V., Bonjean, M., … Maquet, P. (2010). Functional neuroimaging insights into the physiology of human sleep. *Sleep, 33,* 1589–1603.

Datta, S., & Maclean, R. R. (2007). Neurobiological mechanisms for the regulation of mammalian sleep-wake behavior: Reinterpretation of historical evidence and inclusion of contemporary cellular and molecular evidence. *Neuroscience and Biobehavioral Reviews, 31,* 775–824.

Davis, K. F., Parker, K. P., & Montgomery, G. L. (2004). Sleep in infants and young children: Part one: normal sleep. *Journal of Pediatric Health Care, 18,* 65–71.

De Gennaro, L., & Ferrara, M. (2003). Sleep spindles: An overview. *Sleep Medicine Reviews, 7,* 423–440.

Dibner, C., Schibler, U., & Albrecht, U. (2010). The mammalian circadian timing system: Organization and coordination of central and peripheral clocks. *Annual Review of Physiology, 72.* 517–49.

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience, 11,* 114–126.

Dijk, D. J. (2009). Regulation and functional correlates of slow wave sleep. *Journal of Clinical Sleep Medicine, 5,* S6–S15.

Dijk, D. J., & Czeisler, C. A. (1994). Paradoxical timing of the circadian rhythm of sleep propensity serves to consolidate sleep and wakefulness in humans. *Neuroscience Letters, 166,* 63–68.

Ellenbogen, J. M., & Pace-Schott, E. (2012). Drug-induced sleep: Theoretical and practical considerations. *Pfluegers Archive European Journal of Physiology, 463,* 177–186.

Ferrara, M., De Gennaro, L., Curcio, G., Cristiani, R., Corvasce, C., & Bertini, M. (2002). Regional differences of the human sleep electroencephalogram in response to selective slow-wave sleep deprivation. *Cerebral Cortex, 12,* 737–748.

Ferri, R., Cosentino, F. I., Elia, M., Musumeci, S. A., Marinig, R., … Bergonzi, P. (2001). Relationship between Delta, Sigma, Beta, and Gamma EEG bands at REM sleep onset and REM sleep end. *Clinical Neurophysiology, 112,* 2046–2052.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews, 11,* 127–138.

Halassa, M. M. (2011). Thalamocortical dynamics of sleep: Roles of purinergic neuromodulation. *Seminars in Cell and Developmental Biology, 22,* 245–251.

Hernandez, P. J., & Abel, T. (2011). A molecular basis for interactions between sleep and memory. *Sleep Medicine Clinics, 6,* 71–84.

Hobson, J. A. (1989). *Sleep*. New York, NY: Scientific American Library.

Hobson, J. A. (2009). REM sleep and dreaming: Towards a theory of protoconsciousness. *Nature Reviews, 10,* 803–813.

Hobson, J. A., & McCarley, R. W. (1977). The brain as a dream state generator: An activation-synthesis hypothesis of the dream process. *American Journal of Psychiatry, 134,* 1335–1348.

Hobson, J. A., McCarley, R. W., & Wyzinski, P. W. (1975). Sleep cycle oscillation: Reciprocal discharge by two brainstem neuronal groups. *Science, 189,* 55–58.

Hobson, J. A., & Steriade, M. (1986). The neuronal basis of behavioral state control. In F.E. Bloom (Ed.), *Handbook of physiology - the nervous system, Vol. IV*. Bethesda, MD: American Physiological Society.

Iber, C., Ancoli-Israel, E., Chesson, A., & Quan, S. F. (2007). *The AASM manual for the scoring of sleep and associated events: Rules, terminology*

and technical specification. Westchester, IL: American Academy of Sleep Medicine.

Imeri, L., & Opp, M. R. (2009). How (and why) the immune system makes us sleep. *Nature Reviews, 10,* 199–210.

Jones, B. E. (2008). Modulation of cortical activation and behavioral arousal by cholinergic and orexinergic systems. *Annals of the New York Academy of Sciences, 1129,* 26–34.

Kahn, D., Pace-Schott, E. F., & Hobson, J. A. (1997). Consciousness in waking and dreaming: The roles of neuronal oscillation and neuromodulation in determining similarities and differences. *Neuroscience, 78,* 13–38.

Kahn, D., Stickgold, R., Pace-Schott, E. F., & Hobson, J. A. (2000). Dreaming and waking consciousness: A character recognition study. *Journal of Sleep Research, 9,* 317–325.

Kaufmann, C., Wehrle, R., Wetter, T. C., Holsboer, F., Auer, D. P., Pollmacher, T., … Czisch, M. (2006). Brain activation and hypothalamic functional connectivity during human non-rapid eye movement sleep: An EEG/fMRI study. *Brain, 129,* 655–667.

Krueger, J. M., Rector, D. M., & Churchill, L. (2007). Sleep and Cytokines. *Sleep Medicine Clinics, 2,* 161–169.

Krueger, J. M., Rector, D. M., Roy, S., Van Dongen, H. P., Belenky, G., … Panksepp, J. (2008). Sleep as a fundamental property of neuronal assemblies. *Nature Reviews, 9,* 910–919.

Landolt, H., & Dijk, D. (2011). Genetic basis of sleep in healthy humans. In M. H. Kryger, T. Roth & W. C. Dement (Eds.), *Principles and practice of sleep medicine 5th edition*. Philadelphia, PA: Elsevier.

Lavie, P. (1986). Ultrashort sleep-waking schedule. III. "Gates" and "forbidden zones" for sleep. *Electroencephalography and Clinical Neurophysiology, 63,* 414–425.

Lazarus, M., Shen, H. Y., Cherasse, Y., Qu, W. M., Huang, Z. L., Bass, C. E., … Chen, J. F. (2011). Arousal effect of caffeine depends on adenosine A2A receptors in the shell of the nucleus accumbens. *Journal of Neurosciences, 31,* 10067–10075.

Leonard, T. O., & Lydic, R. (1999). Nitric oxide, a diffusible modulator of physiological traits and behavioral states. In B. N. Mallick & S. Inoue (Eds.), *Rapid eye movement sleep* (pp. 167–193). New York, NY: Marcel Dekker.

Leproult, R., & Van Cauter, E. (2010). Role of sleep and sleep loss in hormonal release and metabolism. *Endocrine Development, 17,* 11–21.

Levin, R., & Nielsen, T. A. (2007). Disturbed dreaming, posttraumatic stress disorder, and affect distress: A review and neurocognitive model. *Psychological Bulletin, 133,* 482–528.

Lim, A. S., Lozano, A. M., Moro, E., Hamani, C., Hutchison, W. D., Dostrovsky, J. O., … Murray, B. J. (2007). Characterization of REM-sleep associated ponto-geniculo-occipital waves in the human pons. *Sleep, 30,* 823–827.

Luppi, P. H., Gervasoni, D., Verret, L., Goutagny, R., Peyron, C., Salvert, D., … Fort, P. (2006). Paradoxical (REM) sleep genesis: the switch from an aminergic-cholinergic to a GABAergic-glutamatergic hypothesis. *Journal of Physiology, Paris, 100,* 271–283.

Maquet, P. (2000). Functional neuroimaging of normal human sleep by positron emission tomography. *Journal of Sleep Research, 9,* 207–231.

Marks, G. A., Shaffery, J. P., Oksenberg, A., Speciale, S. G., & Roffwarg, H. P. (1995). A functional role for REM sleep in brain maturation. *Behavioural Brain Research, 69,* 1–11.

Massimini, M., Huber, R., Ferrarelli, F., Hill, S., & Tononi, G. (2004). The sleep slow oscillation as a traveling wave. *Journal of Neurosciences, 24,* 6862–6870.

McCarley, R. W., & Hobson, J. A. (1975). Neuronal excitability modulation over the sleep cycle: A structural and mathematical model. *Science, 189,* 58–60.

McCormick, D. A., & Bal, T. (1997). Sleep and arousal: Thalamocortical mechanisms. *Annual Review of Neuroscience, 20,* 185–215.

Menicucci, D., Piarulli, A., Debarnot, U., D'ascanio, P., Landi, A., … Gemignani, A. (2009). Functional structure of spontaneous sleep slow oscillation activity in humans. *PLoS ONE, 4,* e7601.

Merica, H. (2000). Fast and slow frequency spindles in sleep: Two generators?. *Clinical Neurophysiology, 111,* 1704–1706.

Moore, R. Y., & Card, J. P. (1984). *Noradrenaline-containing Neuron System*. Amsterdam: Elsevier.

Moruzzi, G., & Magoun, H. W. (1949). Brain stem reticular formation and activation of the EEG. *Electroencephalography and Clinical Neurophysiology, 1,* 455–473.

Naidoo, N. (2009). Cellular stress/the unfolded protein response: Relevance to sleep and sleep disorders. *Sleep Medicine Reviews, 13,* 195–204.

Nir, Y., Staba, R. J., Andrillon, T., Vyazovskiy, V. V., Cirelli, C., Fried, I., … Tononi, G. (2011). Regional slow waves and spindles in human sleep. *Neuron, 70,* 153–169.

Nishino, S., Ripley, B., Overeem, S., Lammers, G. J., & Mignot, E. (2000). Hypocretin (orexin) deficiency in human narcolepsy. *Lancet, 355,* 39–40.

Nofzinger, E. A., Buysse, D. J., Germain, A., Carter, C., Luna, B., Price, J. C., … Kupfer, D. J. (2004). Increased activation of anterior paralimbic and executive cortex from waking to rapid eye movement sleep in depression. *Archives of General Psychiatry, 61,* 695–702.

Ohayon, M. M., Carskadon, M. A., Guilleminault, C., & Vitiello, M. V. (2004). Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: Developing normative sleep values across the human lifespan. *Sleep, 27,* 1255–1273.

Ohno, K., & Sakurai, T. (2008). Orexin neuronal circuitry: Role in the regulation of sleep and wakefulness. *Frontiers in Neuroendocrinology, 29,* 70–87.

Pace-Schott, E. F. (2011a). REM sleep and dreaming. In B. N. Mallick, S. R. Pandi-Permual, R. W. McCarley & A. R. Morrison (Eds.), *Rapid eye movement sleep—regulation and function*. Cambridge: Cambridge University Press.

Pace-Schott, E. F. (2011b). The neurobiology of dreaming. In M. H. Kryger, T. Roth & W. C. Dement (Eds.), *Principles and practice of sleep medicine 5th edition*. Philadelphia, PA: Elsevier.

Pace-Schott, E. F., & Hobson, J. A. (2002). The neurobiology of sleep: Genetics, cellular physiology and subcortical networks. *Nature Reviews Neuroscience, 3*(8), 591–605.

Pace-Schott, E. F., & Spencer, R. M. C. (2011). Age-related changes in the cognitive function of sleep. *Progress in Brain Research, 191,* 75–89.

Pannain, S., & Van Cauter, E. (2007). Modulation of endocrine function by sleep-wake homeostasis and circadian rhythmicity. *Sleep Medicine Clinics, 2,* 147–159.

Parmeggiani, P. L. (2003). Thermoregulation and sleep. *Frontiers in Bioscience, 8,* S557–S567.

Paulus, W., & Schomburg, E. D. (2006). Dopamine and the spinal cord in restless legs syndrome: Does spinal cord physiology reveal a basis for augmentation? *Sleep Medicine Reviews, 10,* 185–196.

Peirson, S., & Foster, R. G. (2006). Melanopsin: Another way of signaling light. *Neuron, 49,* 331–339.

Porkka-Heiskanen, T., & Kalinchuk, A. V. (2011). Adenosine, energy metabolism and sleep homeostasis. *Sleep Medicine Reviews, 15,* 123–135.

Prospéro-García, O., Méndez-Díaz, M., Ruiz-Contreras, A. E. A., … Pérez-Morales, M. (2011). Neuropeptides and REM sleep. In B. N. Mallick, S. R. Pandi-Permual, R. W. McCarley & A. R. Morrison (Eds.), *Rapid eye movement sleep—regulation and function*. Cambridge: Cambridge University Press.

Ramirez, J. M., Tryba, A. K., & Peña, F. (2004). Pacemaker neurons and neuronal networks: An integrative view. *Current Opinion in Neurobiology, 14*(6), 665–674.

Rattenborg, N. C., Amlaner, C. J., & Lima, S. L. (2000). Behavioral, neurophysiological and evolutionary perspectives on unihemispheric sleep. *Neuroscience and Biobehavioral Reviews, 24,* 817–842.

Rechtschaffen, A., & Bergmann, B. M. (2002). Sleep deprivation in the rat: An update of the 1989 paper. *Sleep, 25,* 18–24.

Rechtschaffen, A., & Kales, A. (1968). *A manual of standardized terminology techniques and scoring system for sleep stages of human subjects*. Los Angeles, CA: University of California, Brain Information Service/Brain Research Institute.

Samann, P. G., Wehrle, R., Hoehn, D., Spoormaker, V. I., Peters, H., Tully, C., … Czisch, M. (2011). Development of the brain's default mode network from wakefulness to slow wave sleep. *Cerebral Cortex, 21,* 2082–2093.

Saper, C. B., Lu, J., Chou, T. C., & Gooley, J. (2005a). The hypothalamic integrator for circadian rhythms. *Trends in Neurosciences, 28,* 152–157.

Saper, C. B., Chou, T. C., & Scammell, T. E. (2001). The sleep switch: Hypothalamic control of sleep and wakefulness. *Trends in Neuroscience, 24,* 726–731.

Saper, C. B., Fuller, P. M., Pedersen, N. P., Lu, J., & Scammell, T. E. (2010). Sleep state switching. *Neuron, 68,* 1023–1042.

Saper, C. B., Scammell, T. E., & Lu, J. (2005). Hypothalamic regulation of sleep and circadian rhythms. *Nature, 437,* 1257–1263.

Scher, M. S. (2008). Ontogeny of EEG-sleep from neonatal through infancy periods. *Sleep Medicine, 9,* 615–636.

Sherin, J. E., Shiromani, P. J., Mccarley, R. W., & Saper, C. B. (1996). Activation of ventrolateral preoptic neurons during sleep. *Science, 271,* 216–219.

Smith, C. (1995). Sleep states and memory processes. *Behavioural Brain Research, 69,* 137–45.

Solms, M. (1997). *The neuropsychology of dreams: A clinico-anatomical study*. Mahwah, NJ: Lawrence Erlbaum Associates.

Steriade, M. (2000). Corticothalamic resonance, states of vigilance and mentation. *Neuroscience, 101,* 243–276.

Steriade, M. (2006). Grouping of brain rhythms in corticothalamic systems. *Neuroscience, 137,* 1087–1106.

Stickgold, R., Hobson, J. A., Fosse, R., & Fosse, M. (2001). Sleep, learning, and dreams: Off-line memory reprocessing. *Science, 294,* 1052–1057.

Stickgold, R., James, L., & Hobson, J. A. (2000a). Visual discrimination learning requires sleep after training. *Nature Neuroscience, 3,* 1237–1238.

Stickgold, R., Scott, L., Rittenhouse, C., & Hobson, J. A. (1999). Sleep-induced changes in associative memory. *Journal of Cognitive Neuroscience, 11,* 182–193.

Stickgold, R., Whidbee, D., Schirmer, B., Patel, V., & Hobson, J. A. (2000b). Visual discrimination task improvement: A multi-step process occurring during sleep. *Journal of Cognitive Neuroscience, 12,* 246–254.

Szymusiak, R., & McGinty, D. (2008). Hypothalamic regulation of sleep and arousal. *Annals of New York Academy of Science, 1129,* 275–286.

Tononi, G., & Cirelli, C. (2006). Sleep function and synaptic homeostasis. *Sleep Medicine Reviews, 10,* 49–62.

Trinder, J., Kleiman, J., Carrington, M., Smith, S., Breen, S., Tan, N., … Kim, Y. Autonomic activity during human sleep as a function of time and sleep stage. *Journal of Sleep Research, 10,* 253–264.

Van Cauter, E., Holmback, U., Knutson, K., Leproult, R., Miller, A., Nedeltcheva, … Spiegel, K. (2007). Impact of sleep and sleep loss on neuroendocrine and metabolic function. *Hormone Research, 67* (Suppl 1), 2–9.

Van Cauter, E., Leproult, R., & Plat, L. (2000). Age-related changes in slow wave sleep and REM sleep and relationship with growth hormone and cortisol levels in healthy men. *JAMA, 284,* 861–868.

Vertes, R. P., & Kocsis, B. (1997). Brainstem-diencephalo-septohippocampal systems controlling the theta rhythm of the hippocampus. *Neuroscience, 81,* 893–926.

Wagner, U., Gais, S., Haider, H., Verleger, R., & Born, J. (2004). Sleep inspires insight. *Nature, 427,* 352–355.

Walker, M., Brakefield, T., Morgan, A., Hobson, J. A., & Stickgold, R. (2002). Practice with sleep makes perfect: Sleep dependent motor skill learning. *Neuron, 35,* 205–211.

Walker, M. P., & Stickgold, R. (2006). Sleep, memory, and plasticity. *Annual Reviews in Psychology, 57,* 139–166.

Wamsley, E. J., Tucker, M., Payne, J. D., Benavides, J. A., & Stickgold, R. (2010). Dreaming of a learning task is associated with enhanced sleep-dependent memory consolidation. *Current Biology, 20,* 850–855.

Watson, C. J., Baghdoyan, H. A., & Lydic, R. (2011). Neuropharmacology of sleep and wakefulness. *Sleep Medicine Clinics, 5,* 513–528.

Weyand, T. G., Boudreaux, M., & Guido, W. (2001). Burst and tonic response modes in thalamic neurons during sleep and wakefulness. *Journal of Neurophysiology, 85,* 1107–1118.

Suggested Readings

Hobson, J. A. (1988). *The dreaming brain*. New York, NY: Basic Books.

Hobson, J. A., Pace-Schott, E. F., & Stickgold, R. (2000). Dreaming and the brain: Toward a cognitive neuroscience of conscious states. *Behavioral and Brain Sciences, 23,* 793–842. discussion 904–1121.

Jouvet, M. (1999). *The paradox of sleep: The story of dreaming*. Cambridge: MIT Press.

Kryger, M. H., Roth, T., & Dement, W. C. (Eds.) (2011). *Principles and practice of sleep medicine 5th edition*. Philadelphia, PA: Elsevier.

Mallick, B. N., Pandi-Permual, S. R., Mccarley, R. W., & Morrison, A. R. (2011). *Rapid eye movement sleep–regulation and function*. Cambridge: Cambridge University Press.

Steriade, M., & McCarley, R. W. (2005). *Brain control of wakefulness and sleep*. New York, NY: Kluwer Academic/Plenum.

Edward F. Pace-Schott and J. Allan Hobson

This page intentionally left blank

CHAPTER

41

Reward, Motivation, and Addiction

REWARD AND MOTIVATION

Motivation is the process that mediates goal-directed responses to changes in the external or internal environment. Early explanations of motivated behavior focused on reductions of drives or needs, such as for sodium (salt) or water. Such purely homeostatic frameworks for motivation could plausibly account for ingestive and thermoregulatory behaviors and possibly even the drug-seeking behavior generated by the state of drug-withdrawal. Many behaviors, however, are triggered by external incentives which are perceived as appetizing, rewarding, or aversive. Thus, motivated behavior is determined both by biological needs and by learned associations between environmental cues and the value associated with the outcome of particular actions.

Behavioral goals, such as food or a sexual partner, are often not immediately available but must be searched for. Motivated behavior thus comprises both appetitive behavior and consummatory behaviors. *Appetitive behavior*, such as foraging for food or seeking a mate, tends to be adaptive, flexible, and learned through experience. In contrast, *consummatory behavior*, such as ingestion and copulation, tends to be reflexive or stereotyped and acquired early in development. Appetitive behavior occurs in conjunction with endocrine and autonomic responses, such as salivation and insulin secretion (see Chapter 34), which prepare the animal for efficient consummatory behavior.

The complexity of motivated behavior (Fig. 41.1), from the selection of learned voluntary actions to reflexive control of consummatory behavior, involves the coordination of several levels of neural control, including the prefrontal and cingulate cortices, the amygdala, the basal ganglia, the hypothalamus, and neuromodulatory nuclei in the brainstem.

Conditioning and Learning

In the absence of immediate goals, such as food or sex, animals typically use what they have learned from past experiences to predict the outcome of particular actions. Two predominant forms of learning are critical for motivated behavior: *classical conditioning* (i.e., Pavlovian conditioning) and *operant conditioning* (the learned association between goal-directed, instrumental behaviors and their outcomes).

In operant conditioning, outcomes are used to strengthen preceding behaviors via either positive or negative reinforcement. Positive reinforcement increases the likelihood of a particular behavior by having an attractive stimulus, termed a *positive reinforcer*, follow a given behavior. Conversely, negative reinforcement strengthens a behavior by removing an aversive stimulus, or a negative reinforcer, after that behavior is performed. The addition of a positive reinforcer is thought to be rewarding for an animal, whereas the removal of a negative reinforcer is believed to be relieving. (Negative reinforcers are not to be confused with punishers, which serve to decrease, rather than increase, the probability of a response). Reinforcers can be stimuli such as food or drugs which are valuable or aversive in themselves (primary reinforcers), or stimuli that become valuable or aversive by association with primary reinforcers (conditioned reinforcers).

Because motivation depends on the contingency between a particular behavior and a reinforcer, as well as the strength of the reinforcer, researchers can employ operant conditioning paradigms to study an animal's level of motivation. In the self-administration paradigm, animals learn arbitrary instrumental actions, such as a lever press, to gain access to positive reinforcers such as food and drugs (Fig. 41.2). Depending on the value of the reinforcer

© 2013 Elsevier Inc. All rights reserved.

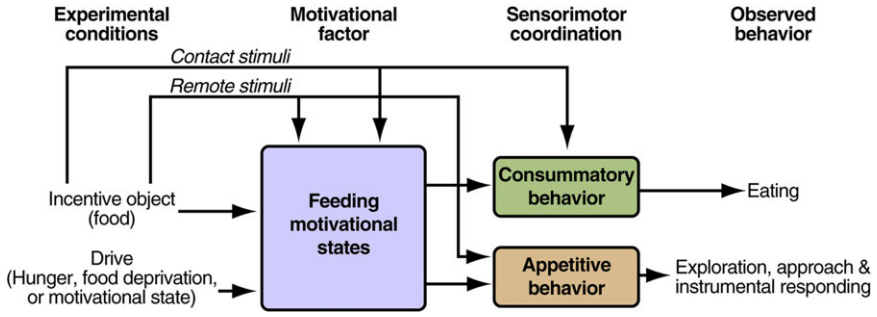


FIGURE 41.1 Theory of incentive-motivation, as applied to feeding. Note that drive (internal stimuli) and food (external stimuli) separately affect central motivation to feed. Also, incentive (i.e., food) and discriminative (i.e., sensorimotor) influences exert their effects separately to control consummatory behavior (when food is contacted) and appetitive behavior (when food remains remote). *Modified with permission from Robbins (1986).*

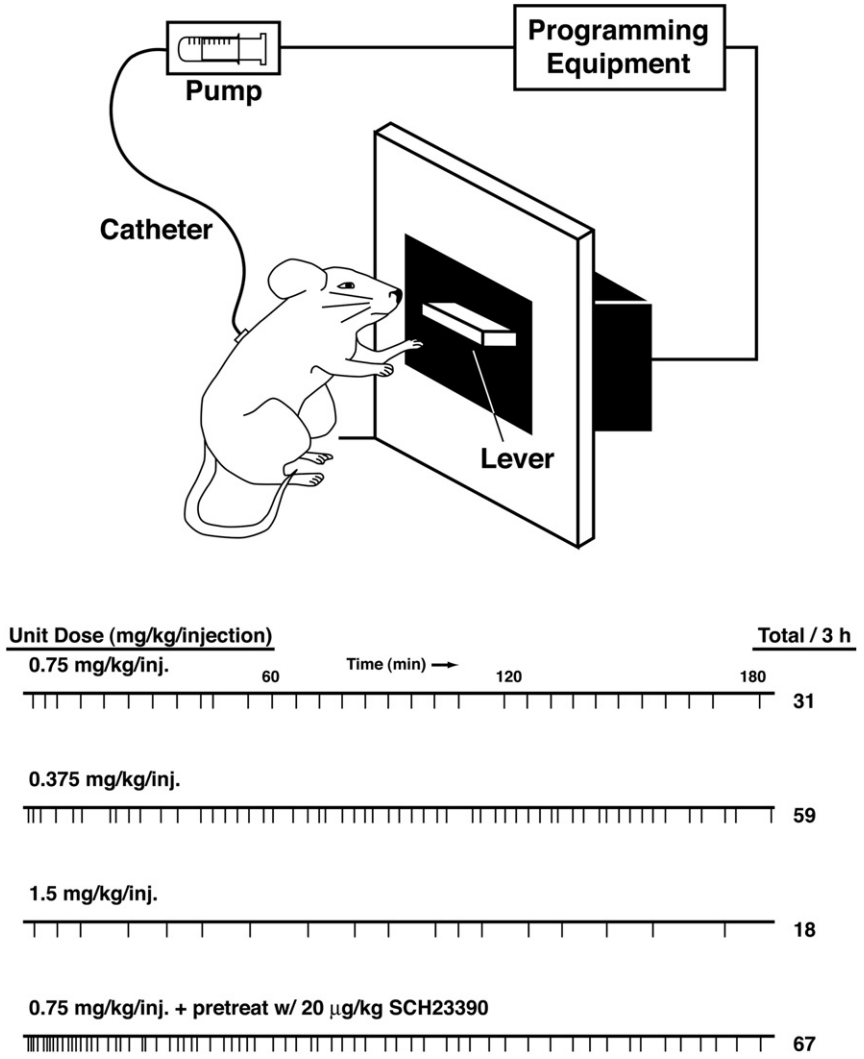


FIGURE 41.2 Intravenous self-administration by rats. (Top) Drawing illustrating the setup for intravenous self-administration of cocaine by rats. (Bottom) Event records for different unit doses of cocaine show a dose-response relationship relating dose of cocaine to number of infusions. Rats implanted with intravenous catheters and trained to self-administer cocaine with limited access (3 h per day) show stable and regular drug intake over each daily session. No obvious tolerance or dependence develops. Rats generally are maintained on a fixed-ratio (FR) schedule of drug infusion, such as FR-1 or FR-5. In an FR-1 schedule, one lever press is required to deliver an intravenous infusion of cocaine; in an FR-5 schedule, five lever presses are required to deliver an infusion of cocaine, and so on. With an FR schedule, rats appear to regulate the amount of drug self-administered. Lowering the dose from the training level of 0.75 mg/kg per injection increases the number of self-administered infusions. Raising the unit dose decreases the number of infusions. Pretreatment with the dopamine receptor antagonist SCH 23390 also increases the number of self-administered infusions. *Reprinted with permission from Caine et al. (1993).*

and the likelihood that their behavior will result in the reinforcer, animals will vary their level of activity (e.g., the amount they lever press). Thus, self-administration is a behavioral paradigm that measures an animal's degree of motivation by assessing its capacity to work for a goal.

Another measure of motivation involves assessing whether animals choose to execute particular behaviors when confronted with alternative choices. A commonly used experimental paradigm is conditioned place preference, in which a choice is made between two locations, only one of which is linked to a positive reinforcer (such as food or a drug). Following several learning trials, a test trial is conducted in the absence of the goal itself. In the test, the preference for that place (as measured by the time spent there) relative to another place without that association is taken as an index of the attractiveness, or incentive properties, of the goal.

The Hypothalamus Plays a Role in Motivation

Inspired by the discovery of brainstem centers for fundamental processes such as respiration, early investigators searched for a motivation center, conceptualized as a neural locus that could integrate internal and external stimuli and thereby coordinate neural circuits to produce adaptive behavioral, endocrine, and autonomic responses. The hypothalamus (Chapter 33) was thought to be ideal for such integrative functions, given its rich blood supply and exposure to cerebrospinal fluid (and thus access to a variety of chemical signals and the capacity also to deliver them), its control over the pituitary, and its neural connections (through brainstem and spinal centers) with visceral afferent information, autonomic and somatic motor outflow.

The hypothalamus is critical for a diverse range of motivated behaviors, including drinking, thermoregulation, aggression, and sex (see Chapter 33). For example, thermoregulatory behavior, such as reflexive shivering, panting, and grooming, is impaired by lesions of the preoptic region of the hypothalamus and is elicited by cooling or warming of this region (Satinoff, 1982). Defensive (affective) aggression and accompanying autonomic changes, such as piloerection, are produced by electrical stimulation of the ventromedial hypothalamus, whereas predatory ("quiet") aggression is produced by electrical stimulation of the lateral hypothalamus (Flynn, Vanegas, Foote, & Edwards, 1970).

The role of the hypothalamus in sexual motivation has been particularly well studied. Some hypothalamic nuclei contain high concentrations of steroid receptors (e.g., for estradiol and testosterone), especially the medial preoptic area and ventromedial hypothalamic nucleus. In male rats, copulatory behavior is abolished by medial preoptic area lesions (Heimer & Larsson, 1966–1967), whereas electrical stimulation of this region elicits copulation. In female rats, lordosis is blocked (and aggression toward males increased) by lesions of the ventromedial hypothalamic nucleus. Lordosis is also induced by electrical stimulation of this region (see Pfaff, 1982). Most of these responses depend on hormones. In ovariectomized females, proceptive ("ear wiggling," hopping, and darting) and receptive (lordosis) behaviors can be elicited by implanting estradiol into the ventromedial hypothalamic nucleus. Similarly, the sexual behavior of castrated males can be restored by implanting testosterone into the medial preoptic area. This hormone dependence underlines the interaction between external (environmental) and internal (plasma steroids) states in inducing motivated responses. Hormones influence the organization of sexually dimorphic brain development (e.g., development of the sexually dimorphic nuclei in the preoptic region), as well as sexually dimorphic behavior. Maternal behavior also is dependent on hormones and is impaired by lesions of the medial preoptic area.

Studies of the role of the hypothalamus in motivation that focus on consummatory, rather than learned instrumental, behavior can yield misleading conclusions. For example, when challenged with thermal stress, rats with preoptic lesions do not exhibit adaptive responses such as reflexive shivering, adjustments in food intake, nest building, or alterations in locomotion. They can, however, achieve thermoregulation by learning to press a lever for hot or cool air and remain motivated to do so (Carlisle, 1969). Similarly, male rats with preoptic area lesions exhibit locomotor excitement in the presence of a receptive female and will press levers to gain access to the female but cannot perform the consummatory behavior of copulation. Male monkeys with preoptic lesions continue to show motivated behavior (e.g., they will masturbate in the presence of females) but are unable to mount and intromit.

This dissociation between consummatory and instrumental behaviors is important because damage to the amygdala is known to produce the opposite pattern of effects—unimpaired mounting but reduced instrumental behavior (Everitt, 1990). These examples show that hypothalamic circuits predominantly control consummatory behaviors rather than learned, goal-directed instrumental behaviors. Behavioral flexibility, which requires learning, depends on the recruitment of additional neural systems, including the amygdala and prefrontal cortex, as described later (see Box 41.1).

Lesions of the Lateral Hypothalamus Disrupt Motivated Behavior and Dopaminergic Axons

Electrolytic lesions of the lateral hypothalamus cause dramatic deficits in food intake which were originally interpreted as a deficit in motivation. However, such lesions destroy far more than the lateral hypothalamic region itself,

BOX 41.1

LOGIC OF METHODOLOGY USED IN BEHAVIORAL NEUROSCIENCE

Theorists (e.g., Hoebel, 1974) have defined principles for identifying the involvement of particular brain regions in motivational (and other behavioral) functions. For example, electrophysiological, neurochemical, or metabolic activity may be observed in a structure, correlated with a behavioral response of interest, which suggests an involvement of that structure with the function under study. However, this correlative relationship does not establish whether the functioning of that region is causally important for the behavior. This issue may be addressed by an intervention that disrupts (impairs or improves) the function under study. For example, if the function is lost following selective lesioning, this suggests that the brain structure is implicated in that function. Similarly, if the function is enhanced following a boosting of neurotransmission in that region (e.g., via mild levels of electrical or chemical stimulation through implanted electrodes or cannulae), this is similarly converging evidence of a causative role. However, it is important to realize that the use of these techniques is each associated with interpretative problems, and their effects must be critically appraised. For example, the designation of the lateral hypothalamic region as a "feeding center" was based on the following: (1) electrolytic lesions produced profound aphagia (abstention from eating); (2) electrical stimulation of the brain through electrodes implanted in the lateral hypothalamus induced eating in sated rats (Hoebel, 1974); (3) infusions of the neurotransmitter norepinephrine into the hypothalamus also induced eating in sated rats (Leibowitz, 1980); and (4) electrophysiological recording showed that when monkeys were deprived of food, individual neurons in the lateral hypothalamus were sensitive to the sight and taste of food (Rolls, 1985). While many of these experimental observations have remained valid, subsequent evidence has suggested that the lateral hypothalamus has a much more specific role in feeding motivation, as a result of using improved techniques (see main text). Moreover, many of the effects associated with diminishing and increasing activity within the lateral hypothalamus have been shown subsequently to depend upon coincidental influences on axons within the medial forebrain bundle which merely pass through the lateral hypothalamus (e.g., between midbrain and striatum or between prefrontal cortex and brainstem) and had little to do functionally with the lateral hypothalamus itself. Only through the use of several neurobiological techniques providing converging evidence can the relationships between brain and behavior become more certain.

since they also interrupt ascending and descending fibers in the medial forebrain bundle, including dopaminergic axons (Fig. 41.3). Four major systems that are responsible for conveying dopamine to specific regions of the forebrain have been identified. The diencephalic incertohypothalamic dopamine system provides dopaminergic terminals to the hypothalamus, including the medial preoptic area responsible for sexual behavior (Pfaus, 2009). The tuberoinfundibular dopamine system controls hormone release from the anterior pituitary gland. The mesolimbic dopamine system conveys dopamine from neurons in the midbrain ventral tegmental area (VTA) to "limbic" regions of the forebrain, including the nucleus accumbens of the ventral striatum, amygdala, and medial prefrontal cortex. The nigrostriatal dopamine system originates in the substantia nigra and provides dopaminergic innervation to the caudate and putamen of the dorsal striatum.

Consistent with a critical role for dopamine in motivated behavior, injections of the catecholamine-selective neurotoxin 6-hydroxydopamine (6-OHDA) into the vicinity of the lateral hypothalamus reproduced many of the effects of electrolytic lesions (Ungerstedt, 1971a). The injected rats suffered large depletions (>90%) of dopamine from the striatum. They were cataleptic and had difficulty initiating head, limb, axial, and oral movements. Further experiments showed that the syndrome could be produced by injection of 6-OHDA at any of several points along the nigrostriatal dopamine pathway. Thus, dopamine depletion at sites far from the lateral hypothalamus reproduced much of the lateral hypothalamic syndrome. 6-OHDA-induced lesions of the mesolimbic pathway minimally affected ingestion but did disrupt other behaviors that can accompany feeding, including locomotion and exploration (Koob, Riley, Smith, & Robbins, 1978). These findings demonstrate that dopamine depletion impairs both consummatory and appetitive behaviors. Direct axonal projections from orexin neurons in the hypothalamus to dopamine neurons in the VTA provide a functional link between the homeostatic and incentive motivational components of ingestive behavior.

Dopamine Is Critical for Motivated Behavior

What is the nature of the contribution by dopaminergic neurons to motivation? Clues to answering this question were originally provided by experiments in which rats received unilateral 6-OHDA lesions of the nigrostriatal

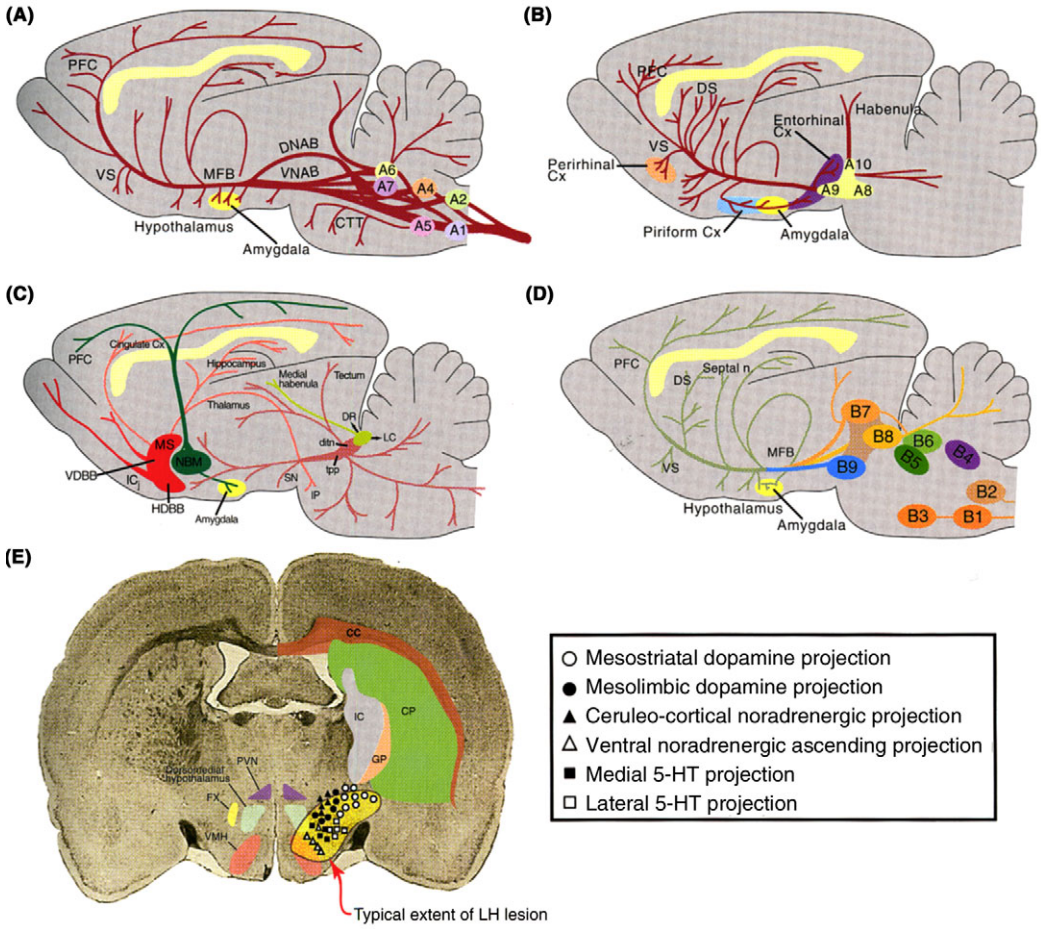


FIGURE 41.3 Ascending monoamine neurotransmitter systems. The figure shows schematic sagittal (A–D) and coronal (E) sections through the lateral hypothalamus of a rat brain. (A) Origin and distribution of central noradrenergic pathways. Note noradrenergic cell groups A1–A7, including the locus coeruleus (A6). DNAB, dorsal noradrenergic ascending bundle; VNAB, ventral noradrenergic ascending bundle. (B) Origin and distribution of central dopamine pathways. Note dopaminergic cell groups A8–A10. (C) Origin and distribution of central cholinergic pathways. Note rostral cell groups. NBM, nucleus basalis magnocellularis (Meynert in primates); MS, medial septum; VDBB, vertical limb nucleus of the diagonal band of Broca; HDBB, horizontal limb nucleus of the diagonal band of Broca. (D) Origin and distribution of central serotonergic pathways. Note cell groups in the raphe nucleus, B4–B9. MFB, medial forebrain bundle; PFC, prefrontal cortex; VS, ventral striatum; DS, dorsal striatum. *Based on Robbins and Everitt (1995).* (E) Schematic coronal section through the rat brain at the level of the ventromedial hypothalamus. LH lesions could disrupt various ascending monoaminergic fibers. CC, corpus callosum; CP, caudate-putamen; DMH, dorsomedial hypothalamus; IC, internal capsule; GP, globus pallidus; PVH, paraventricular hypothalamus; LH, lateral hypothalamus; FX, fornix; VMH, ventromedial hypothalamus. *Modified from Robbins (1986), with permission.*

dopamine circuit. The rats rotated toward the side of the lesion following treatment with a dopamine-releasing drug, such as D-amphetamine (Ungerstedt, 1971b). Presumably, this rotation occurred because more dopamine is released on the intact side than on the lesioned side of the brain and because any output pathway from the striatum influences movements in the opposite side of the body. Of even greater interest, the direct dopamine receptor agonist apomorphine caused the rats to turn away from the side of the lesion, an effect explained by actions at supersensative dopamine receptors on the lesioned side.

Unilateral lesions of either the lateral hypothalamus or the nigrostriatal dopamine pathway produce polymodal sensorimotor neglect, measured by responses to contralateral somatic, auditory, and olfactory stimuli (Marshall & Teitelbaum, 1977). Although this neglect was originally attributed to sensory inattention, further analysis suggests that the fundamental deficit is a failure in initiating responses rather than in attention (see Box 41.2). Animals with lesions of the nigrostriatal dopamine system retain the capacity for movement but have great difficulty in initiating movement. Similarly, patients suffering from Parkinson's disease have severe deficits in nigrostriatal dopamine with concommitant impairments in initiating movement. These considerations suggest that nigrostriatal dopamine is critical for activating or invigorating behavioral responses.

Studies with molecular and genetic techniques have also implicated dopamine in behavioral activation (Palmiter, 2009). Dopamine production relies on the synthetic

BOX 41.2

SEPARATION OF SENSORY AND RESPONSE FACTORS IN THE DOPAMINERGIC SYNDROME OF NEGLECT

Sensory and motor aspects of the neglect syndrome were separated in the following way (Carli et al., 1985). Rats were trained to detect brief flashes of light presented to either side of the head. In one case, they were trained to respond where the light was. In the alternate case, rats were trained to respond in a location away from the side of the light. Rats with dopamine depleted from one side of the striatum could not respond properly toward the contralateral side, whether or not the stimulus was presented on that side. Measurement of reaction time showed that the problem was in the initiation, rather than completion, of the movement. Consistent with these data, dopamine-depleted rats could detect a contralateral stimulus and respond to it with an unlateralized response. Thus, their deficit was in the initiation of movement in response to stimuli rather than in the detection of the stimuli.

enzyme tyrosine hydroxylase, which is also critical for the production of norepinephrine. Genetic deletion of tyrosine hydroxylase specifically in dopaminergic neurons produces severe impairments in mice; they become hypoactive and aphagic, and they starve by 4 weeks of age. Daily treatment with L-DOPA restores activity and feeding, demonstrating the critical role of dopamine for goal-directed behaviors necessary for survival. Restoration of dopamine synthesis specifically to the dorsal striatum via stereotactic injections of a recombinant virus is sufficient to rescue deficits in feeding and the initiation of locomotory behavior, underscoring the critical role of dopaminergic innervation of striatal neurons for motivated behaviors.

Dopaminergic Neuron Firing Is Related to Incentive and Reward

Electrophysiological recordings from midbrain dopaminergic neurons in animals have provided a framework for understanding the role of dopamine in motivated behavior (Bromberg-Martin & Hikosaka, 2010; Schultz, Dayan, & Montague, 1997). Midbrain dopaminergic neurons fire tonically at low rates of about 5 Hz in awake animals. The resulting tonic release of dopamine is critical for the ongoing operation of downstream circuits. Stimuli that are either directly rewarding (such as juice) or predict rewards (such as a cue which has been paired with a subsequent reward) evoke phasic increases in firing which elevate dopamine concentration at synaptic terminals for a period of several seconds.

A critical insight into the function of dopamine neurons was revealed by the pioneering studies of Wolfram Schultz and colleagues (Schultz et al., 1997). Dopaminergic neurons in the VTA originally fire in response to the presentation of a reward, but if a sensory cue is provided which predicts the reward, dopamine neurons fire in response to the cue instead of the reward (Fig. 41.4). Remarkably, if a fully predicted reward fails to occur, dopamine neurons decrease rather than increase their firing rate. These and other findings have led to the hypothesis that dopamine neuronal firing encodes "reward prediction errors." The dynamic changes in dopamine neuron firing as learning proceeds match quantitative predictions from reinforcement learning and have inspired researchers to borrow principles from economics to guide neuroscientific studies (Glimcher, 2011).

Motivated behaviors can be directed toward rewarding stimuli and some dopamine neurons appear to encode the motivational salience of stimuli, responding to stimuli that are either rewarding or aversive (Matsumoto & Hikosaka, 2009). Signals encoding aversive stimuli (or the lack of rewarding stimuli) are thought to be conveyed to dopamine neurons indirectly from a nucleus in the diencephalon called the habenula (Hikosaka et al., 2010). Neurons in the habenula decrease their firing rate in response to rewarding stimuli and increase their firing rates in response to aversive stimuli and when expected rewards are absent. Dopaminergic neurons in the substantia nigra which fire in response to both rewarding and aversive stimuli are hypothesized to convey motivational signals to the superior colliculus and frontal cortex, brain regions responsible for orienting attention toward and evaluating potentially salient stimuli, as well as to the striatum.

Electrical Stimulation of the Brain Can Be a Positive Reinforcer

Olds and Milner (1954) discovered that rats learned new responses when their learning was positively reinforced by electrical stimulation of the brain, and the sites that supported robust electrical brain self-stimulation followed the trajectory of the medial forebrain bundle. That finding affirmed the importance of incentive factors in instrumental behavior. The best natural parallel appeared to be behavior in response to large incentives by animals in low states of drive (e.g., chocolate milk in nondeprived

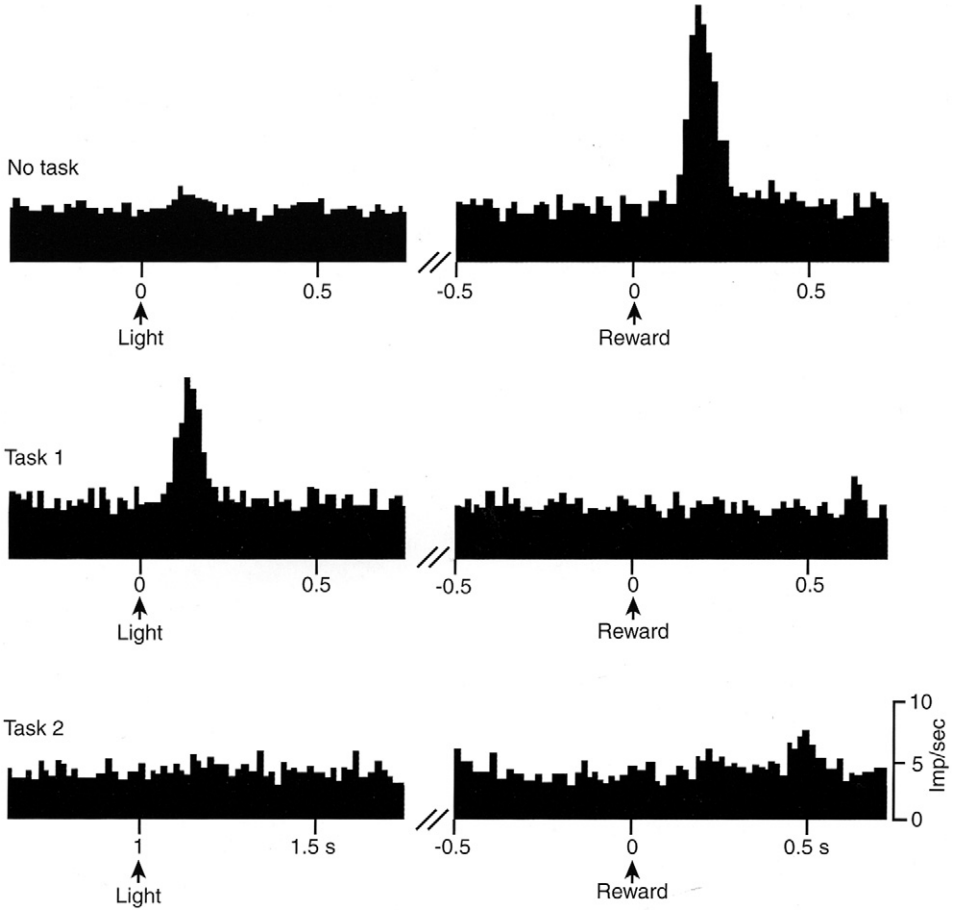


FIGURE 41.4 Responses of a dopamine neuron to unpredicted primary reward (top) and the transfer of this response to progressively earlier reward-predicting conditioned stimuli with training (middle). The bottom record shows a control baseline task when the reward is predicted by an earlier stimulus and not the light. *From Schultz et al. (1995), with permission.*

rats). Indeed, measurement of the strength of reinforcers in microamperes (Box 41.3 and Fig. 41.5) promised a pragmatic, as well as objective, new approach to the psychophysics of an operational hedonism. However, perhaps its greatest contribution was to enable the anatomical mapping of the central reward or reinforcement mechanisms of the brain (Phillips, 1984).

Catecholamine Hypothesis of Rewarding Brain Stimulation

Subsequent experiments indicated that animals can learn to engage in intracranial self-stimulation (ICSS) in several other brain regions (e.g., the locus coeruleus, midbrain central gray, cerebellum, trigeminal motor nucleus, substantia nigra, caudate nucleus, and nucleus accumbens) (Phillips, 1984). Although the diversity of these sites has not been fully elucidated, many of them are known to contain catecholamine-containing neurons, and while such neurons can express either norepinephrine or dopamine, most pharmacologic data can be interpreted in terms of dopaminergic mechanisms (Box 41.3 and Fig. 41.5). In particular, very high rates of ICSS were found in the ventral tegmental area, ventral striatum (including the nucleus accumbens to which dopamine neurons project), all key structures of the medial forebrain bundle. However, of several neural circuits that may subserve rewarding brain stimulation, only some are influenced directly by dopamine. The difficulty in interpreting the role of striatal dopamine in ICSS lies in distinguishing motor from reward effects (Box 41.3). The phenomenon of human drug dependence and the discovery in the early 1960s of drug self-administration in animals opened the way to more precisely analyze the neurochemical basis of reinforcement.

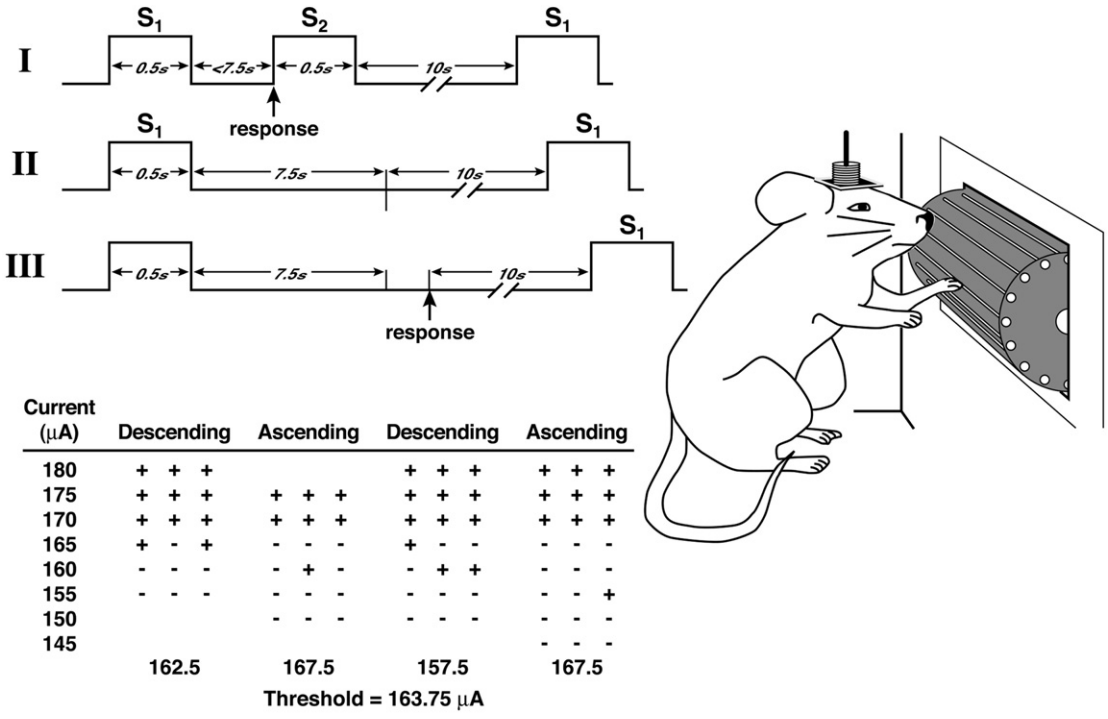
The Mesolimbic Dopamine System Plays a Key Role in Incentive Motivation

Wise (1982) proposed that natural reinforcers are perceived as rewards because they increase activity of the mesolimbic dopamine system. If this is correct, then the mesolimbic dopamine system should be important for behavioral responses to natural rewards, such as food and sex. In fact, depletion of dopamine in the nucleus accumbens does not impair consummatory behavior in rats (unlike dopamine depletion from the caudate-putamen), but it does reduce incentive-motivational responses. For example, proceptive behaviors (such as "ear wiggling,"

BOX 41.3

PSYCHOPHYSICS OF HEDONISM AS MEASURED BY BRAIN STIMULATION REWARD

Electrical stimulation of the brain can be used as positive reinforcement of learned behaviors. An apparatus is set up so that when an animal presses a lever, its brain is stimulated. Alternatively, a threshold procedure can be used where a lever or wheel response is required. An advantage of using electrical stimulation to measure motivation is that the strength of the reward can be titrated easily by adjusting the amount of current delivered for each response. If the current is adjusted too low (i.e., below the threshold for reinforcement), then the animals will not press the lever. As current is increased, the relationship between the rate of lever pressing and the strength of current typically follows a sigmoidal curve called rate-intensity function. In the threshold procedure, a series of discrete trials are presented at different current intensities, and the responses within a certain cutoff period are recorded. Current levels are lowered and raised to capture the threshold current (Fig. 41.2). Pharmacologic manipulation of catecholamine neurotransmitter systems can lower thresholds (e.g., treatment with amphetamines or cocaine) or raise thresholds (e.g., treatment with the dopamine receptor antagonist pimozide). Such changes mean the amount of electrical current required to support lever presses varies, while the latency to respond is unchanged. Thus, these drugs affect reward processes, rather than motor function. The threshold of reinforcement also can be measured by allowing rats to adjust the strength of the current. In this setup, a process is employed in which rats press lever one to produce ICSS, but each lever-one press also reduces the intensity of the current. To reset current intensity, rats must press lever two. Such "self-titration" experiments yield conclusions similar to those based on threshold and rate-intensity functions.



Current (μA)	Descending	Ascending	Descending	Ascending
180	+ + +		+ + +	+ + +
175	+ + +	+ + +	+ + +	+ + +
170	+ + +	+ + +	+ + +	+ + +
165	+ - +	- - -	+ - -	- - -
160	- - -	- + -	- + +	- - -
155	- - -	- - -	- - -	- - +
150		- - -	- - -	- - -
145				- - -
	162.5	167.5	157.5	167.5

Threshold = 163.75 μA

FIGURE 41.5 Intracranial self-stimulation threshold procedure. Panels I, II, and III illustrate the timing of events during three hypothetical discrete trials. Panel I shows a trial during which the rat responded within the 7.5 s following the delivery of the noncontingent stimulus (positive response). Panel II shows a trial during which the animal did not respond (negative response). Panel III shows a trial during which the animal responded during the intertrial interval (negative response). For demonstration purposes, the intertrial interval was set at 10 s. In reality, the interresponse interval had an average duration of 10 s and ranged from 7.5 to 12.5 s. The table depicts a hypothetical session and demonstrates how thresholds were defined for the four individual series. The threshold of the session is the mean of the four series' thresholds. *Taken with permission from Markou and Koob (1992).*

hopping, and darting, all of which serve to solicit the attention of the male) are reduced in female rats with dopamine depletions. In addition, depletion of dopamine from the nucleus accumbens decreases locomotor excitement of hungry rats in the presence of food (Koob, Riley, Smith, & Robbins, 1978). Intriguingly, consummatory behavior, including eating of high incentive foods, appears to be mediated by distinct mechanisms within the nucleus accumbens that depend on opioid receptors (Hnasko, Sotak, & Palmiter, 2005; Kelley and Berridge, 2002).

Mesolimbic dopamine also may control behavior motivated by reward. For example, it was shown in studies by Robbins and Taylor that when a previously neutral light became a predictor of the occurrence of a primary reinforcer (water), rats learned a new behavior to turn on the light. Injection of amphetamine into the nucleus accumbens (but not the caudate) increased the frequency of the behavior, suggesting that amphetamine enhanced the motivational properties of this reward-related stimulus (i.e., the light, which was related to the reward, water). Amphetamine itself was well known to suppress eating and drinking; thus, the increased frequency of behavior was not due to increased thirst, for example. Furthermore, amphetamine did not increase the frequency of responding to a randomly paired stimulus. Finally, amphetamine-induced facilitation was completely blocked by dopamine depletion in the nucleus accumbens (but not by dopamine depletion in the caudate). These results indicated that the mesolimbic dopamine system enhances the motivational properties of stimuli predictive of natural rewards (Robbins, Taylor, Cador, & Everitt, 1989). Another way of describing this activational role of the dopamine system is that it confers "incentive salience" (Berridge & Robinson, 1998).

The Amygdala Also Has a Role in Appetitive Conditioning and Motivation

Growing evidence implicates the amygdala in evaluating both positive and negative stimuli (Morrison & Salzman, 2010). Excitotoxic, axon-sparing lesions of the amygdala diminish the capacity of stimuli associated with reward to motivate behavior. Such lesions also impair responses that are rewarded with access to female rats. However, mating *per se* is unaffected, in contrast with the effect of medial preoptic area lesions, (see above and Everitt, Cardinal, Hall, Parkinson, Robbins, 2000). This separation of control of consummatory and learned aspects of sexual behavior indicates that motivated sequences of behavior are constructed through the coordination of neural systems that are at least partly independent. Lesions of the basolateral amygdala also impair conditioned place preferences. In contrast, selective lesions of the central nucleus of the amygdala impair the learning of Pavlovian conditioned approach responses.

Motivated Behavior Translated to Habits

The ventral striatum, including the nucleus accumbens, receives afferents from the basolateral amygdala, hippocampal formation, and prefrontal cortex, and projects to the lateral hypothalamus and ventral pallidum (Fig. 41.6). Output from the ventral pallidum is routed in several different ways, including to brainstem motor regions and back to the prefrontal cortex through the mediodorsal thalamus. This circuit (prefrontal cortex-ventral striatum-ventral pallidum) appears to fit the requirements for "a neural mechanism by which motivation gets translated into action" (Mogenson, Jones, & Yim, 1980).

As described earlier, the basal and lateral amygdala convey associative information about stimuli that predict the occurrence of reinforcers. A link between the amygdala and ventral striatum is important in the translation of emotion (and motivational effects of stimuli) to behavioral output (Everitt et al., 2000). Glutamatergic inputs from the amgydala to ventral striatum work together with the ascending dopamine system to determine the output of ventral striatal γ-aminobutyric acid (GABA)-ergic medium spiny neurons that project to the globus pallidus (ventral pallidum) (see Chapter 30).

The nucleus accumbens is a heterogeneous structure. Its "shell" region is considered by some to be part of an "extended amygdala" encompassing the central nucleus of the amygdala and bed nucleus of the stria terminalis. The basolateral amygdala projects predominantly to the nucleus accumbens core subregion and specific parts of the dorsal striatum. These connections are paralleled by other cortico-striatal "loops" (see Chapter 30) (e.g., between different parts of the neocortex that project topographically to different sectors of the dorsal striatum). Moreover, there appears to be the possibility of cross-talk between these systems, arising from anatomical connections that potentially allow communication from the amygdala-ventral striatal circuitry to the cortico-dorsal striatal loops (Haber, Fudge, & McFarland, 2000). The ventral striatum has been linked to the initial acquisition of drug-seeking and drug-taking behavior and the subsequent cascade of loop circuitry involving the dorsal striatum to the gradual acquisition of stimulus-response habits that help to give the habits their "compulsive" quality (Belin, Jonkman, Dickinson, Robbins, & Everitt 2009; Everitt and Robbins, 2005).

ADDICTION

Epidemiology of Drug Abuse

Drug abuse is one of the world's major public health problems. For example, statistics from the 2009 *National*

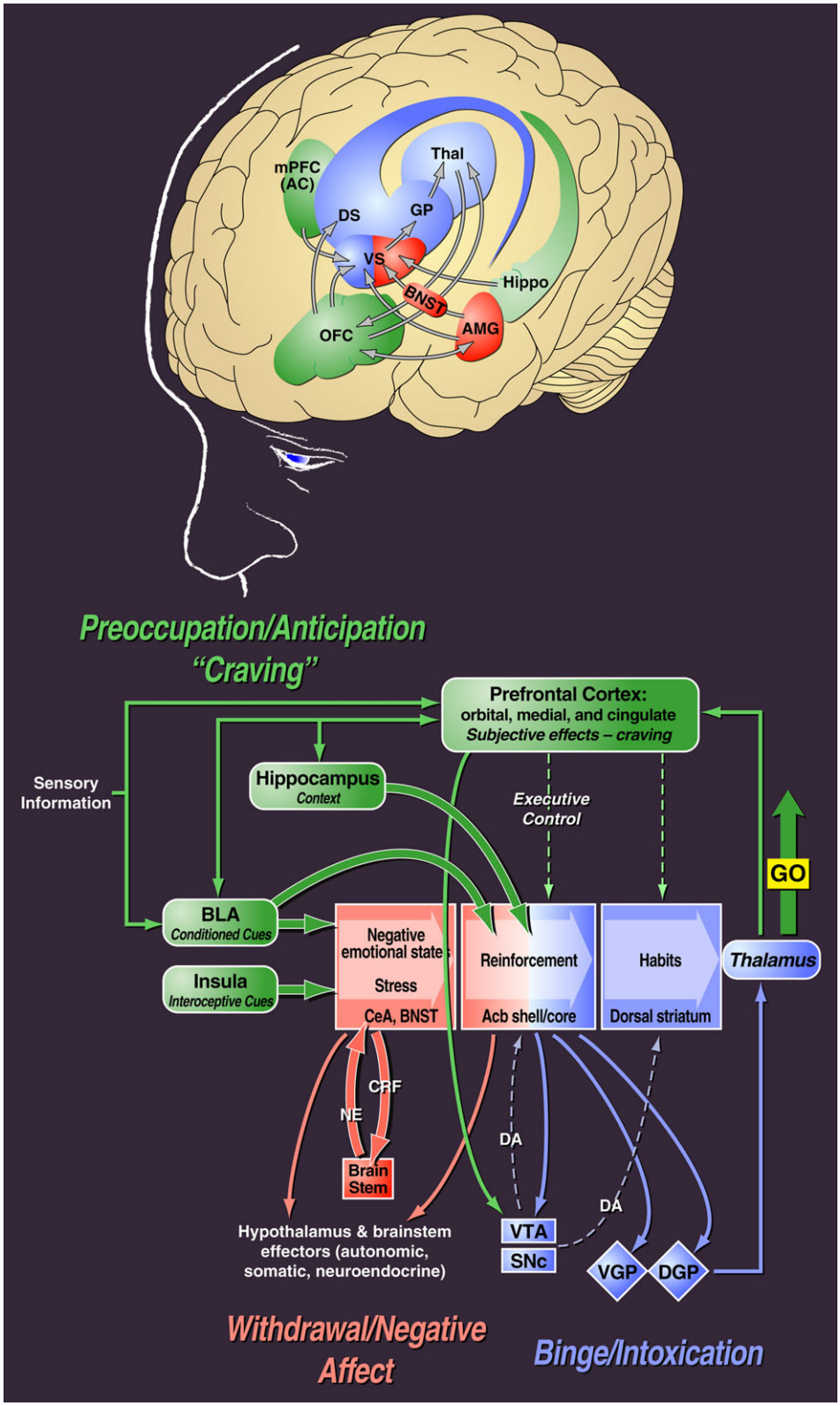


FIGURE 41.6 Neurocircuitry schematic illustrating the combination of neuroadaptations in the brain circuitry for the three stages of the addiction cycle that drive drug-seeking behavior in the addicted state. Note the activation of the ventral striatum/dorsal striatum/extended amygdala driven by cues via hippocampal and basolateral amygdala and stress via the insula. The frontal cortex system is compromised, producing deficits in executive function and contributing to the incentive salience of drugs compared to natural reinforcers. Dopamine systems are compromised, and brain stress systems such as corticotropin-releasing factor (CRF) are activated to reset further the salience of drugs and drug-related stimuli in the context of an aversive dysphoric state. *Modified with permission from Koob and Volkow (2010) and based on Everitt and Robbins (1995).*

Household Survey on Drug Abuse of the National Institute on Drug Abuse of the United States reveal that approximately 21.8 million Americans aged 12 and older were past-month users of illicit drugs, and 22.5 million could be classified with substance dependence or abuse. In addition, 15.4 million individuals were dependent on or abused alcohol, and 69.7 million were current smokers (Substance Abuse and Mental Health Services Administration, 2010). The cost of drug abuse to society is staggering. Total economic costs of alcohol abuse per year in the United States is listed as $166.5 billion dollars, smoking $157 billion dollars (Centers for Disease Control and Prevention, 2004), and illicit drugs $193 billion dollars (National Drug Intelligence Center, 2011). This totals over $500 billion dollars in lost productivity due to illness or death, health care expenditures, motor vehicle crashes, and social welfare, among others (Stein, 2001).

Addiction Cycle

Drug addiction, also known as substance dependence, is a chronically relapsing disorder characterized by (1) compulsion to seek and take the drug, (2) loss of control in limiting intake, and (3) emergence of a negative emotional state (e.g., dysphoria, anxiety, irritability) when access to the drug is prevented (Koob & Le Moal, 1997). *Addiction* and *substance dependence* (as currently defined by the *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition; American Psychiatric Association, 1994) will be used interchangeably throughout this text and refer to a final stage of a usage process that moves from drug use to addiction. Clinically, the occasional but limited use of a drug with the *potential* for abuse or dependence is distinct from escalated drug use and the emergence of a chronic drug-dependent state (Table 41.1). An important goal of current neurobiological research is to understand

TABLE 41.1 DSM-IV and ICD-10 Diagnostic Criteria for Alcohol and Drug Dependence

	DSM-IV	ICD-10
Clustering criterion	A. A maladaptive pattern of substance use, leading to clinically significant impairment or distress as manifested by three or more of the following occurring at any time in the same 12-month period:	A. Three or more of the following have been experienced or exhibited at some time during the previous year:
Tolerance	1. Need for markedly increased amounts of a substance to achieve intoxication or desired effect; or markedly diminished effect with continued use of the same amount of the substance.	1. Evidence of tolerance, such that increased doses are required in order to achieve effects originally produced by lower doses.
Withdrawal	2. The characteristic withdrawal syndrome for a substance or use of a substance (or a closely related substance) to relieve or avoid withdrawal symptoms.	2. A physiological withdrawal state when substance use has ceased or been reduced as evidenced by: the characteristic substance withdrawal syndrome, or use of substance (or a closely related substance) to relieve or avoid withdrawal symptoms.
Impaired control	3. Persistent desire or one or more unsuccessful efforts to cut down or control substance use. 4. Substance used in larger amounts or over a longer period than the person intended.	3. Difficulties in controlling substance use in terms of onset, termination, or levels of use.
Neglect of activities	5. Important social, occupational, or recreational activities given up or reduced because of substance use.	4. Progressive neglect of alternative pleasures or interests in favor of substance use; or
Time spent	6. A great deal of time spent in activities necessary to obtain, to use, or to recover from the effects of substance used.	a great deal of time spent in activities necessary to obtain, to use, or to recover from the effects of substance use.
Inability to fulfill roles	None	None
Hazardous use	None	None
Continued use despite problems	7. Continued substance use despite knowledge of having a persistent or recurrent physical or psychological problem that is likely to be caused or exacerbated by use.	5. Continued substance use despite clear evidence of overtly harmful physical or psychological consequences.
Compulsive use	None	6. A strong desire or sense of compulsion to use substance.
Duration criterion	B. No duration criterion separately specified. However, several dependence criteria must occur repeatedly as specified by duration qualifiers associated with criteria (e.g., "often," "persistent," "continued").	B. No duration criterion separately specified.
Criterion for subtyping dependence	*With physiological dependence*: Evidence of tolerance or withdrawal (i.e., any of items A-1 or A-2 above are present). *Without physiological dependence*: No evidence of tolerance or withdrawal (i.e., none of items A-1 or A-2 above are present).	None

the neuropharmacological and plastic mechanisms within specific neural circuits that mediate the transition from occasional, controlled drug use to the loss of behavioral control over drug-seeking and drug-taking that defines chronic addiction.

Addiction has been conceptualized as a chronic relapsing disorder with roots both in impulsivity and compulsivity and neurobiological mechanisms that change as an individual moves from one domain to the other (Hyman, Malenka, & Nestler, 2006). Subjects with impulse control disorders experience an increasing sense of tension or arousal before committing an impulsive act; pleasure, gratification, or relief at the time of committing the act; and finally, regret, self-reproach, or guilt following the act. In contrast, individuals with compulsive disorders experience anxiety and stress before committing a compulsive repetitive behavior and then relief from the stress by performing the compulsive behavior. In addiction, drug-taking behavior progresses from impulsivity to compulsivity in a three-stage cycle: *binge/intoxication*, *withdrawal/negative affect*, and *preoccupation/anticipation*. As individuals move from an impulsive to a compulsive disorder, the drive for the drug-taking behavior shifts from positive to negative reinforcement (Fig. 41.7). Impulsivity persists, but compulsivity develops (Box 41.4). Impulsivity in general is an aberration of reward anticipation processes, mainly in the ventral striatum (plus possibly response preparation processes in the dorsal striatum), whereas compulsivity is in part the product of habit structures mediated by the dorsal striatum (plus possibly ventral striatal Pavlovian overmotivation; Everitt & Robbins, 2005). Compulsivity also involves the additional negative reinforcement recruited from sensitization of the brain stress systems (Koob & Le Moal, 2005).

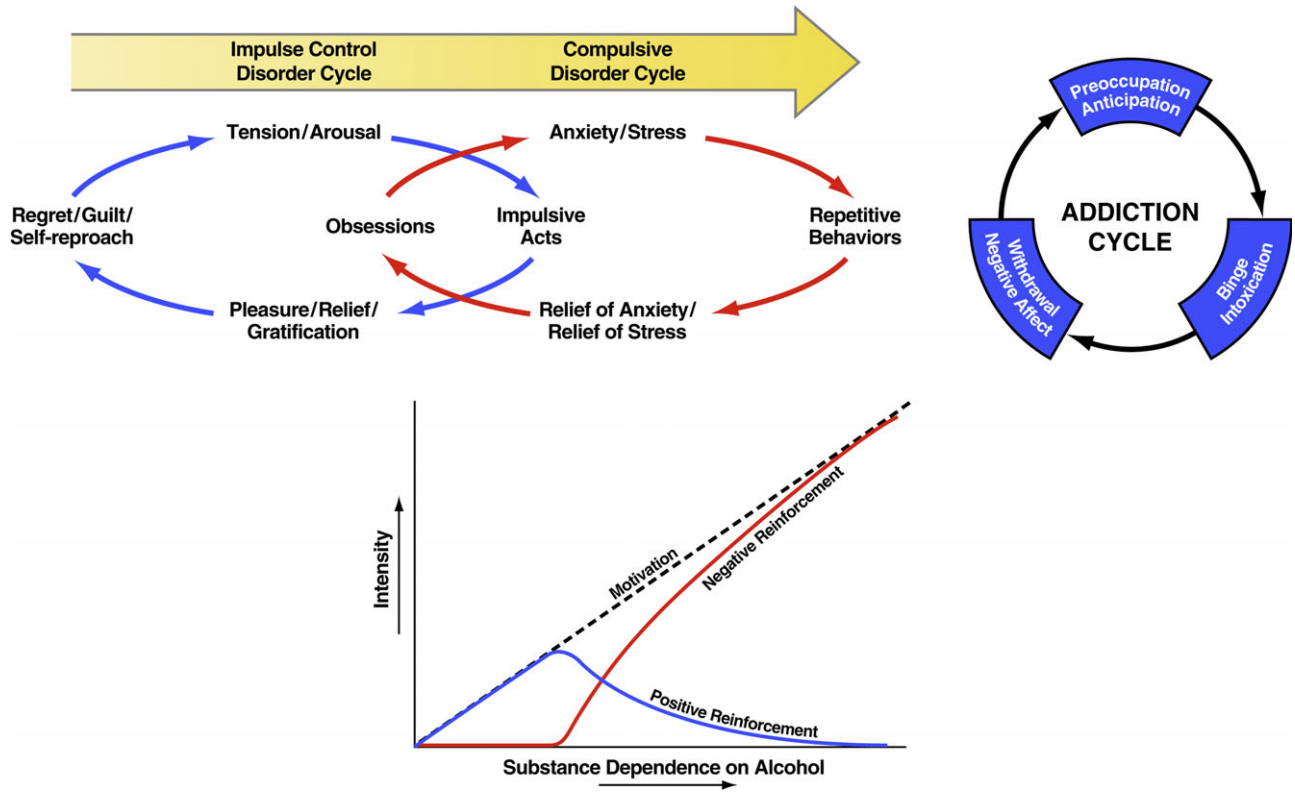


FIGURE 41.7 (Top) Diagram showing the stages of impulse control disorder and compulsive disorder cycles related to the sources of reinforcement. In impulse control disorders, an increasing tension and arousal occurs before the impulsive act, with pleasure, gratification, or relief during the act. Following the act, there may or may not be regret or guilt. In compulsive disorders, there are recurrent and persistent thoughts (obsessions) that cause marked anxiety and stress, followed by repetitive behaviors (compulsions) that are aimed at preventing or reducing distress (American Psychiatric Association, 1994). Positive reinforcement (pleasure/gratification) is more closely associated with impulse control disorders. Negative reinforcement (relief of anxiety or relief of stress) is more closely associated with compulsive disorders. (Bottom) Change in the relative contribution of positive and negative reinforcement constructs during the development of substance dependence on alcohol. *Taken with permission from Koob, (2011).* (Top right) Collapsing the cycles of impulsivity and compulsivity results in the addiction cycle, conceptualized as three major components: *preoccupation/anticipation*, *binge/intoxication*, and *withdrawal/negative affect. Taken with permission from Koob (2008).*

BOX 41.4

IMPULSIVITY AND COMPULSIVITY

Impulsivity can be defined behaviorally as "actions which are poorly conceived, prematurely expressed, unduly risky, or inappropriate to the situation and that often result in undesirable consequences" (Durana and Barnes, 1993). However, *compulsivity* can be captured by a modification of the definition of impulsivity: actions inappropriate to the situation which persist, have no obvious relationship to the overall goal, and which often result in undesirable consequences (e.g., operationally defined as responding for a drug or alcohol in the face of adverse consequences (Wolffgramm & Heyne, 1995), responding in the face of punishment (Vanderschuren & Everitt, 2004), perseveration in reversal learning (Walker, Robbins, & Roberts, 2009), or responding for a drug or alcohol on a progressive-ratio schedule of reinforcement (Walker, Rasmussen, Raskind, Koob, 2008)).

Much of the recent progress in understanding the neurobiology of addiction has derived from the study of animal models of addiction on specific drugs such as opioids, stimulants, and alcohol. While no animal model of addiction fully emulates the human condition, animal models do permit the investigation of specific elements of the process of drug addiction. Such elements can be defined by psychological constructs such as positive and negative reinforcement and different stages of the addiction cycle.

Neurobiological Substrates of Drug Reward

Positive Reinforcing Effects of Drugs Are Mediated by Multiple Systems Converging on Common Targets in the Basal Forebrain

As noted above, the medial forebrain bundle is involved in brain stimulation reward and natural rewards, and work in the neurobiology of addiction has led to an understanding of the neurochemical and neuroanatomical components of this system. The principal components of this system include the ventral tegmental area, basal forebrain (nucleus accumbens, olfactory tubercle, frontal cortex, and amygdala), and dopamine connection between the VTA and basal forebrain, called the mesolimbic dopamine system. Additional components are the opioid peptide, GABA, serotonin (5-hydroxytryptamine; 5-HT), and endocannabinoid systems that interact with the ventral tegmental area and the basal forebrain (Koob, 1992) (Fig. 41.8). The functional role of each component of the basal forebrain reward system is discussed in the following sections. The actions of several groups of drugs with respect to this system, including indirect sympathomimetics (cocaine and amphetamine), opioids (heroin), nicotine, sedative hypnotics (alcohol, barbiturates, and benzodiazepines), and Δ^9-tetrahydrocannabinol (THC) also will be discussed.

The Positive Reinforcing Effects of Cocaine and Other Indirect Sympathomimetics Depend Critically on the Mesolimbic Dopamine System

Amphetamine and cocaine are psychomotor stimulants that, in humans, have behavioral effects such as suppressing hunger and fatigue and inducing euphoria. In animals, these drugs increase locomotor activity, decrease food intake, stimulate operant behavior, enhance conditioned responding, and produce place preferences. Psychomotor stimulants also decrease thresholds for reinforcing brain stimulation and act as reinforcers for drug self-administration (Koob, 1992) (Fig. 41.8).

Psychomotor stimulants with a high potential for abuse have effects that lead to increases in the availability of monoamine neurotransmitters at synapses. By blocking reuptake or enhancing release, cocaine and amphetamine increase the synaptic availability of dopamine, norepinephrine, and serotonin. However, the acute reinforcing effects of these drugs depend critically on dopamine. Studies of intravenous self-administration have provided the most direct evidence implicating dopamine and, more specifically, the mesolimbic dopamine system in the reinforcing actions of cocaine and amphetamines. Low doses of dopamine receptor antagonists injected either systemically or centrally into the nucleus accumbens, amygdala, or bed nucleus of the stria terminalis block cocaine and amphetamine self-administration in rats. A specific role for the mesolimbic dopamine system in the reinforcing properties of cocaine and amphetamine was deduced from the observation that neurotoxin-induced lesions of the terminal regions of the mesolimbic dopamine system in the nucleus accumbens produce a significant and long-lasting decrease in self-administration of cocaine and amphetamine over days (Koob, 1992). 6-OHDA-induced lesions of the dopamine terminals in the nucleus accumbens even decrease the amount of work an animal will perform for cocaine without affecting work for many other reinforcers, such as food.

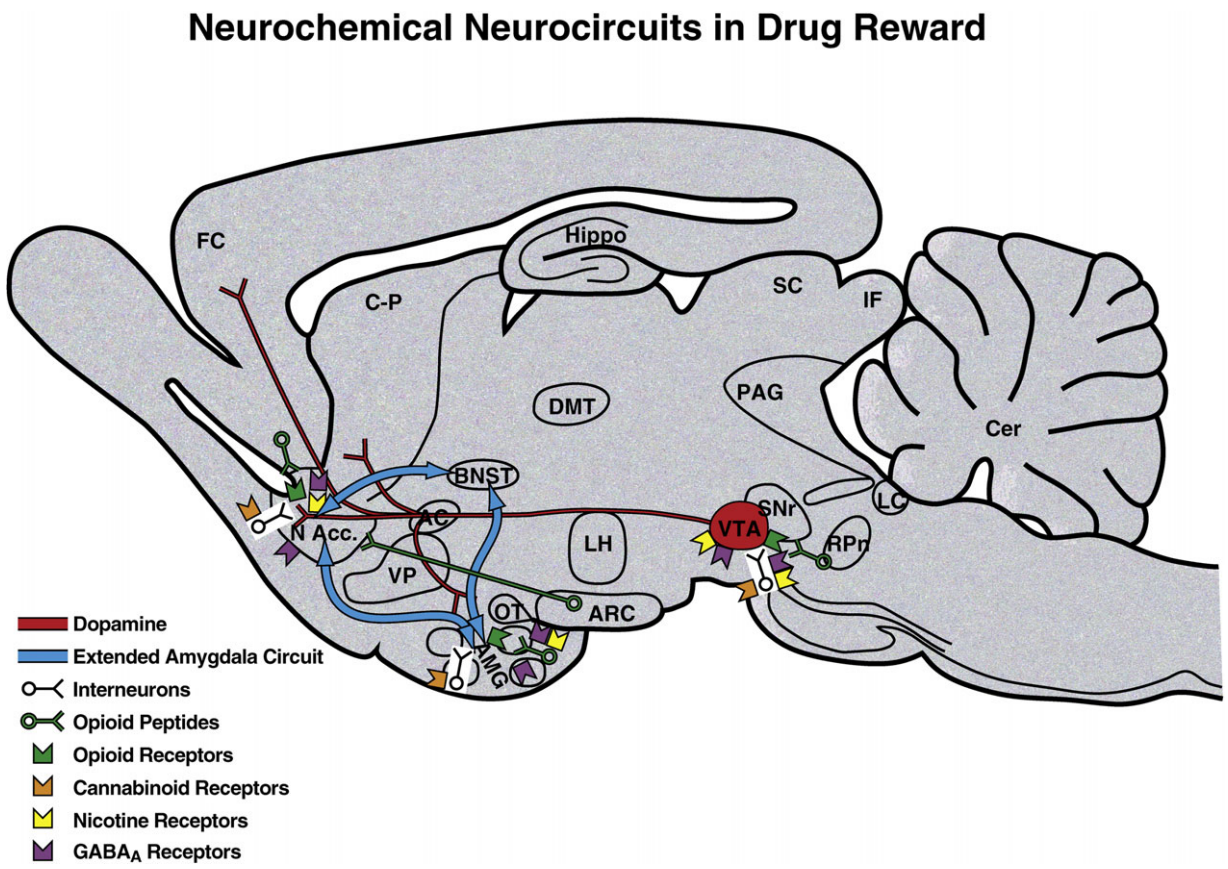


FIGURE 41.8 Sagittal section through a representative rodent brain illustrating the pathways and receptor systems implicated in the acute reinforcing actions of drugs of abuse. Cocaine and amphetamines activate the release of dopamine in the nucleus accumbens and amygdala via direct actions on dopamine terminals. Opioids activate opioid receptors in the ventral tegmental area, nucleus accumbens, and amygdala via direct actions on interneurons. Opioids facilitate the release of dopamine in the nucleus accumbens via an action either in the ventral tegmental area or the nucleus accumbens (but also are hypothesized to activate elements independent of the dopamine system). Alcohol activates γ-aminobutyric acid-A ($GABA_A$) receptors in the ventral tegmental area, nucleus accumbens, and amygdala via either direct actions at the $GABA_A$ receptor or through indirect release of GABA. Alcohol is hypothesized to facilitate the release of opioid peptides in the ventral tegmental area, nucleus accumbens, and central nucleus of the amygdala. Alcohol facilitates the release of dopamine in the nucleus accumbens via an action either in the ventral tegmental area or the nucleus accumbens. Nicotine activates nicotinic acetylcholine receptors in the ventral tegmental area, nucleus accumbens, and amygdala, either directly or indirectly, via actions on interneurons. Nicotine also may activate opioid peptide release in the nucleus accumbens or amygdala independent of the dopamine system. Cannabinoids activate CB_1 receptors in the ventral tegmental area, nucleus accumbens, and amygdala via direct actions on interneurons. Cannabinoids facilitate the release of dopamine in the nucleus accumbens via an action either in the ventral tegmental area or the nucleus accumbens, but also are hypothesized to activate elements independent of the dopamine system. Endogenous cannabinoids may interact with postsynaptic elements in the nucleus accumbens involving dopamine and/or opioid peptide systems. The blue arrows represent the interactions within the extended amygdala system hypothesized to have a key role in drug reinforcement. AC, anterior commissure; AMG, amygdala; ARC, arcuate nucleus; BNST, bed nucleus of the stria terminalis; Cer, cerebellum; C-P, caudate-putamen; DMT, dorsomedial thalamus; FC, frontal cortex; Hippo, hippocampus; IF, inferior colliculus; LC, locus coeruleus; LH, lateral hypothalamus; N Acc., nucleus accumbens; OT, olfactory tract; PAG, periaqueductal gray; RPn, reticular pontine nucleus; SC, superior colliculus; SNr, substantia nigra pars reticulata; VP, ventral pallidum; VTA, ventral tegmental area. *Taken with permission from Koob (2005).*

Neural Substrates of Sensitization Involve Changes in the Mesolimbic and Nigrostriatal Dopamine Systems

Under certain circumstances, repeated administration of stimulants and opioids results in increasingly strong motor activation, and these effects have been hypothesized to contribute to the development of the neuroadaptations associated with addiction, particularly those involving initial use and/or reinstatement of drug-taking following abstinence. This sensitization to the activating effects of stimulants involves activation of the mesolimbic and nigrostriatal dopamine system (Robinson & Berridge, 1993). Repeated microinjections of amphetamine into the somatodendritic region of the ventral tegmental dopamine cells, at doses that do not cause behavioral activation, are sufficient to sensitize the dopamine cells to later systemic injections of amphetamine. This sensitization can be manifested as increased locomotor activation or

increased stereotyped-like behavior (see above). Thus, changes in the activity of these dopamine cells are sufficient to produce sensitization.

Multiple mechanisms for sensitization have been proposed, and a common theme is a time-dependent chain of adaptations that ultimately leads to long-lasting changes in the function of the mesolimbic dopamine system (Robinson & Berridge, 1993). For example, repeated administration of cocaine produces a decrease in the sensitivity of impulse-regulating somatodendritic dopamine D_2 receptors on dopamine neurons (termed *autoreceptors* because the neuron is itself responding to the transmitter it releases). Because these autoreceptors normally exert an inhibitory influence on the activity of dopaminergic neurons, this change in sensitivity could translate into enhanced dopamine availability with subsequent injections of cocaine (Robinson & Berridge, 1993). However, the decrease in D_2 receptor sensitivity lasts only 4–8 days, and behavioral sensitization can last for weeks; thus, other mechanisms also must be involved. *N*-methyl-D-aspartate (NMDA) glutamate receptor antagonists selectively block the development of sensitization to psychomotor stimulants, suggesting a role for brain glutamate systems and specific glutamate receptors in sensitization (Robinson & Berridge, 1993). In addition, an increased surface-to-intracellular ratio of glutamate-1 receptors (GluR1) has been observed 21 days after the last injection of cocaine, suggesting a slowly developing redistribution of AMPA receptors to the surface of nucleus accumbens neurons, particularly in those lacking GluR2 subunits (Boudreau & Wolf, 2005).

Stressors can cause sensitization to stimulant drugs, and both the stress axis and the extrahypothalamic corticotropin-releasing factor (CRF) system appear to be important in stress-induced sensitization (Koob, 1999). The hypothalamic-pituitary-adrenal stress axis (see Chapter 33) also may play an important role in sensitization. Sensitization may be hypothesized to play a key role in initial sensitivity to drugs and contribute to the high incentive salience attributed to drugs during the course of addiction. However, sensitization of the reward systems is not readily observed in individuals meeting the criteria for addiction, rather a profound tolerance and negative emotional state during withdrawal are the clinical norm.

Neurobiological Substrates for the Acute Reinforcing Effects of Opioids Involve Opioid Peptide Systems

Much like psychostimulants, opioid drugs such as heroin are readily self-administered intravenously by animals, and with limited access opioids do not produce dependence (Schuster & Thompson, 1969). Advances in opioid pharmacology, such as the identification of high specific binding for opioids in the brain and discovery of endogenous opioid peptides, provided insight into the brain systems responsible for the reinforcing effects of these drugs (Koob, 1992) (Box 41.5 and Figs. 41.7 and 41.10). The reinforcing actions of heroin and morphine appear to be mediated largely by the μ opioid receptor subtype. Antagonists at the μ receptor produce dose-dependent decreases in heroin reinforcement, and μ receptor knockout mice show loss of morphine-induced reward and loss of morphine-induced analgesia. Intracerebral opioid receptor antagonists block heroin self-administration in nondependent rats when these drugs are injected into the VTA or the region of the nucleus accumbens. Microinjections of opioids into the VTA also lower thresholds for ICSS and produce robust place preferences, and these effects appear to be dependent on mesolimbic dopamine function (Di Chiara & North, 1992). Rats will self-administer opioid peptides in the region of the nucleus accumbens. However, heroin self-administration is not blocked by dopamine receptor antagonists administered at doses that block cocaine self-administration or by large neurotoxin-induced lesions of the mesolimbic dopamine system. These

BOX 41.5

OPIOID RECEPTORS

In 1978, Professors Solomon H. Snyder, John Hughes, and Hans W. Kosterlitz were awarded the Lasker award for their combined discovery of opioid receptors in the brain and for the even more amazing discovery of the existence of endogenous ligands for opioid receptors. The endogenous brain compounds were peptides, called enkephalins and endorphins, which shared all the physiologic properties of opioids. These observations opened a rich field of inquiry into the mechanism of action of opioid drugs, and these findings began a new outlook on neuropeptide neurotransmitters, an area of intense research. First, many new neuropeptides have been discovered with brain receptors potentially amenable to nonpeptidergic agonists and antagonists. Second, because neuropeptides are contained within neurons that also are more conventional amino acid or amine transmitters, the possibility of multiple mediators for any specific synapse had to be considered.

results demonstrate that neural elements in the regions of the VTA and nucleus accumbens are responsible for the reinforcing properties of opioids, suggesting both dopamine-dependent and dopamine-independent opiate actions.

Nicotine Activates Mesolimbic Dopamine

Nicotine has anti-fatigue, stimulant-like, and anti-anxiety effects and appears to improve cognitive performance in animals and humans. It has direct reinforcing actions, as measured by intravenous self-administration in animals and humans. Nicotine is a direct agonist at nicotinic acetylcholine receptors (see Chapter 8) and appears to activate nicotinic acetylcholine receptors in the mesolimbic dopamine system both at the level of the VTA and nucleus accumbens (Figs. 41.7 and 41.10). Knockout of the $\alpha 4$ and $\beta 2$ nicotinic acetylcholine receptor subunit blocks nicotine self-administration (Picciotto et al., 1998). Blockade of dopaminergic systems with antagonists or lesions to the mesolimbic dopamine system can also block nicotine self-administration, suggesting a key role for dopamine in the acute reinforcing actions of nicotine.

Alcohol Has Multiple Neurochemical Substrates within Brain Reinforcement Systems

Alcohol (i.e., ethanol), barbiturates, and benzodiazepines all have measurable sedative–hypnotic actions, including euphoria, disinhibition, anxiety reduction, sedation, and hypnosis. All of these drugs also have anxiolytic (or antianxiety) effects, reflected in conflict situations as a reduction of behavior suppressed by punishment. The neurobiological basis of the anxiolytic properties of these sedative–hypnotic drugs has provided clues to their reinforcing properties and their abuse potential. For example, GABAergic receptor antagonists reverse many of the behavioral effects of alcohol, an observation that has led to the hypothesis that GABA has a role in alcohol's intoxicating effects. Further support for a role of brain GABA in alcohol reinforcement is the observation that potent antagonists of GABA receptor function decrease alcohol reinforcement, and one particularly affected brain site is the central nucleus of the amygdala (Koob, 1992). Activation of opioid peptide systems has been implicated in alcohol reinforcement by numerous reports that the opiate antagonists naloxone and naltrexone reduce alcohol self-administration in several animal models (Tabakoff & Hoffman, 1992). The brain site most sensitive to opioid receptor antagonists are the VTA and central nucleus of the amygdala (Figs. 41.7 and 41.10).

Dopamine receptor antagonists also reduce lever pressing for alcohol in nondeprived rats (Koob, 1992), and extracellular dopamine levels increase in nondependent rats orally self-administering low doses of alcohol (Koob & Le Moal, 2001). Low doses of acetaldehyde are also self-administered by animals, and increasing evidence suggests that acetaldehyde may mediate some of ethanol's effects on the mesolimbic dopamine system (Rodd et al., 2005). However, virtually complete destruction of dopamine terminals in the nucleus accumbens by 6-OHDA failed to alter voluntary responding for alcohol (Tabakoff & Hoffman, 1992). Combined with the pharmacologic data discussed earlier, these results suggest that although mesolimbic dopamine transmission may be associated with important aspects of alcohol reinforcement, it is not critical for the reinforcing properties of alcohol, and multiple neurotransmitters combine to orchestrate the reward profile of this drug (Engel, et al. 1987). For example, alcohol in a physiologic dose range may antagonize the actions of glutamate. In addition, increases and decreases in synaptic availability of serotonin (e.g., via blockade of serotonin reuptake or via administration of selective serotonin receptor antagonists) reduce the voluntary intake of alcohol.

Δ^9-Tetrahydrocannabinol Has Effects Similar to Other Drugs of Abuse

THC shares effects in animal models of drug reinforcement similar to those of other drugs of abuse. Upon acute administration, THC decreases ICSS reward thresholds and produces a place preference in rats. It maintains self-administration behavior in squirrel monkeys, and a synthetic THC analog is self-administered intravenously in mice. THC binds to the cannabinoid CB_1 receptor that is distributed widely throughout the brain, particularly in the striatum, and is thought to mediate some of the actions of the endogenous endocannabinoids. THC self-administration is blocked by both CB_1 and opioid receptor antagonists (Justinova, Tanda, Munzar, & Goldberg, 2004; Justinova, Solinas, Tanda, Redhi, & Goldberg, 2005) (Figs. 41.7 and 41.10).

Neurobiological Substrates for the Withdrawal/Negative Affect State of Drug Dependence

Substance dependence is defined as compulsive, uncontrollable drug use; however, the etiology of that compulsive use is multidimensional. Although positive reinforcement is clearly necessary for the development of drug use, it may fall short in explaining the development of compulsive use. For example, tolerance or apparent tolerance develops to drug reward during chronic use, with positive reinforcement either being absent or subsumed by negative reinforcement. Indeed, this issue raises the question of what factors distinguish drug *use* from *abuse* and *dependence* or addiction.

Negative Affect Is a Common Result of Withdrawal from Chronic Administration of Drugs of Abuse

Withdrawal signs associated with the cessation of chronic drug administration usually are characterized by responses that are opposite to the initial effects of the drug. The physical signs (or somatic signs) of withdrawal are typically drug-specific. However, all drugs of abuse produce rewarding effects, and withdrawal commonly is associated with subjective symptoms of dysphoria, negative affect, and anxiety (Koob & Le Moal, 2001). In the case of opioid withdrawal, subjective symptoms such as dysphoria, anxiety, craving for drugs, and malaise are accompanied by both physical symptoms of extreme discomfort, including pupillary dilation, hot and cold flashes, goose bumps, and a flu-like state. In the case of alcohol withdrawal, humans show tremor and increases in heartbeat rate, blood pressure, and body temperature, as well as subjective signs, including dysphoria, anxiety, insomnia, and malaise. Indeed, whereas few physical signs of withdrawal usually are observed during cocaine withdrawal in humans, cessation of cocaine use often is characterized by severe depressive symptoms combined with irritability, anxiety, and anhedonia lasting several hours to several days (i.e., the "crash") and may be a motivating factor in the maintenance of the cocaine dependence cycle. Withdrawal from nicotine is characterized by anxiety, negative affect, fatigue, irritability, sleep disturbances, and an intense craving for cigarettes. Withdrawal from THC resembles a combination of sedative–hypnotic or opioid withdrawal, with irritability, restlessness, hot flashes, sweating, disturbances in appetite and sleep, and low-level dysphoria. Withdrawal from benzodiazepines resembles that of alcohol withdrawal but generally is more prolonged in onset and less intense.

Chronic Drug Administration Compromises the Brain Reward Systems

In animal studies, withdrawal from chronic administration of stimulants has been studied using ICSS after repeated administration of cocaine or amphetamine over several weeks or 12–48 h of cocaine self-administration. Withdrawal from prolonged self-administration of cocaine in rats increases ICSS reward thresholds. Here, an increase in reward thresholds has been hypothesized to reflect a dysphoric-like state observed in humans. The increase in threshold dose is time-dependent and is an effect opposite to that of acute cocaine (Koob & Le Moal, 2001). Similar effects have been observed during withdrawal from chronic amphetamine and methamphetamine administration. Withdrawal from chronic alcohol, opioids, nicotine, and THC also produces pronounced increases in ICSS reward thresholds, suggesting that brain reward dysregulation is a common element of acute withdrawal from drugs of abuse (Table 41.2).

TABLE 41.2 Drug Effects on Thresholds of Rewarding Brain Stimulation

Drug Class	Acute Administration	Withdrawal from Chronic Treatment
Psychostimulants (cocaine, amphetamines)	↓	↑
Opioids (morphine, heroin)	↓	↑
Nicotine	↓	↑
Sedative–hypnotics (ethanol)	↓	↑
Cannainoids	↓	↑

Neural Circuitry Mediating Withdrawal/Negative Reinforcement

A brain system that mediates aversion was identified early on by a combination of anatomical and behavioral experiments. For example, mapping studies identified sites eliciting escape from brain stimulation, sites at which electrical brain stimulation elicits defensive responses, and sites at which fear or pain naturally evokes aggressive responses. One component of the neural circuitry that mediates aversion includes connections from the medial habenula to interpeduncular nucleus, which in turn projects to brainstem serotonergic neurons as well as projections from the lateral habenula to the rostral tegmental nucleus, which provides inhibitory inputs to dopaminergic neurons in the VTA and substantia nigra.

Aversion May Be Mediated by Opponent Motivational Processes

Animal learning theory has led to the concept of opponent motivational processes (one appetitive, one aversive). An influential model proposed which reinforcers produce affective and hedonic effects (*a-processes*) which are opposed by *b-processes* of opposite affective sign in a simple, dynamic control system for affect (Fig. 41.9) (Solomon & Corbit, 1974). The *a-process* follows the reinforcing event with short latency and then decays. Its size decreases with repeated presentation of the reinforcer. The *b-process* has a longer latency of onset, reaches its maximum only after repeated trials, and decays very slowly. These components of emotional responses are exhibited in response to many reinforcers, including not only drugs such as heroin, but also natural rewards that also may lead to different forms of "dependence," such as attachment of offspring to a mother. So, for example, either heroin or the mother elicits a hedonic *a-process*, which is followed by a rebound negative or aversive *b-process* when the heroin wears off (or the mother leaves). These effects may

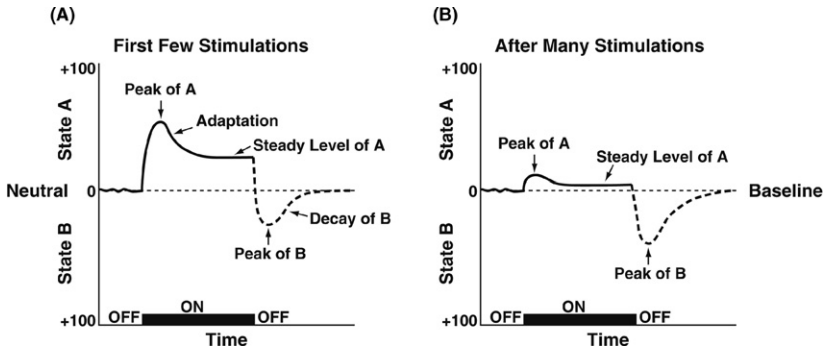


FIGURE 41.9 (A) The standard pattern of affective dynamics produced by a relatively novel unconditioned stimulus (nondependent drug intake). (B) The standard pattern of affective dynamics produced by a familiar, frequently repeated unconditioned stimulus (dependent drug intake). First, an unconditional arousing stimulus triggers a primary affective process, termed the *a-process*. An unconditional reaction that translates the intensity, quality, and duration of the stimulus (for example, the first injection of a drug). Second, as a consequence of the *a-process*, the *b-process* is evoked after a short delay, an opponent process. The two responses are consequently and temporarily linked (*a* triggers *b*) but are hypothesized to depend on different neurobiological mechanisms. The *b-process* has a longer latency but more inertia, slower recruitment, and more sluggish decay. At a given moment, the pattern of affect will be the algebraic sum of these opposite influences and the dynamics reveal, with the passage of time, the net product of the opponent process. *Taken with permission from Solomon (1980).*

get stronger with repeated experiences of heroin (or the mother at a critical period of development), leading to a form of dependence in each case. The opponent process also can work in reverse. For example, if tolerance develops for an initially aversive experience, then a rebound hedonic response can occur when this experience stops. Indeed, this has been suggested partly to explain the "pleasure" of jogging.

How might these opponent processes be mediated? The neuroadaptive mechanisms that reflect changes in reward function have been hypothesized to be alterations of the same neurotransmitters implicated in the acute reinforcing effects of drugs (Koob & Bloom, 1988). Examples of such homeostatic, "within-system" adaptations include decreases in dopaminergic and serotonergic transmission in the nucleus accumbens during drug withdrawal as measured by *in vivo* microdialysis (Parsons, Koob, & Weiss, 1995; Weiss, Markou, Lorang, & Koob, 1992), increased sensitivity of opioid receptor transduction mechanisms in the nucleus accumbens during opioid withdrawal (Shaw-Lutchman et al., 2002), and decreased GABAergic and increased NMDA glutamatergic transmission during ethanol withdrawal (Fitzgerald & Nestler, 1995; Roberts, Cole, & Koob, 1996; Weiss et al., 1996). However, recruitment of other neurotransmitter systems in the adaptive responses to drugs of abuse may involve neurotransmitter systems not linked to the acute reinforcing effects of the drug, or a "between-system" neuroadaptation (Koob & Bloom, 1988). An example of a between-system neuroadaptation is the activation of brain stress systems such as CRF and dysregulation of other stress-mediating neurotransmitter systems, such as dynorphin, during withdrawal (Koob & Le Moal, 2005) (see below).

Neurochemical Adaptation in Reward Neurotransmitters

Dopamine and Serotonin Systems Show Neuroadaptive Changes with Substance Dependence

Monoamines, including dopamine and serotonin, have been implicated in the reinforcing actions of each of the drugs discussed: psychostimulants, opioids, nicotine, alcohol, and THC. One way of assessing the prolonged effects of drug self-administration is to measure extracellular levels of a neurotransmitter during ongoing behaviors, such as intravenous self-administration of a drug or withdrawal from the drug after a period of chronic self-administration. Such measurements can be made by using the technique of *in vivo* microdialysis, which involves implanting into a specific site in the brain a guide cannula through which a probe with a semipermeable membrane at the tip perfuses and collects the neurotransmitter in extracellular fluid. Several studies using this technique have shown that extracellular dopamine and serotonin levels in the nucleus accumbens decrease during acute withdrawal from cocaine, opioids, and alcohol, an effect opposite to that of acute drug administration. Consistent with these results, many studies have shown that acute withdrawal from all drugs of abuse results in decreased firing of VTA dopamine neurons (Melis, Spiga, & Diana, 2005).

Brain Stress Systems: Corticotropin-Releasing Factor, Noradrenergic, and Dynorphin-κ Opioid Systems Show Neuroadaptive Changes with Substance Dependence

Anxiety, stress, and dysphoria are common elements of substance dependence and acute withdrawal from drugs

of abuse in humans, and withdrawal from drugs of abuse produces stress-like behavior in animals (Sarnyai, Shaham, & Heinrichs, 2001). Stress can be defined simply as *any challenge to or alteration in psychological homeostatic processes*. Animal studies exploring the role of brain stress systems in the actions of alcohol, opioids, cocaine, nicotine, and THC have provided evidence of a neurochemical basis for the stress associated with abstinence following chronic drug administration. One major player is CRF. Not only is CRF a major hypothalamic releasing factor controlling the classic glucocorticoid stress response, but it also appears to have a neurotropic role in the central nervous system, modulating behavioral responses to stress. CRF itself produces stress-related behaviors, and CRF antagonists reverse a number of behavioral responses to stress. Rats treated repeatedly with drugs such as alcohol, opioids, and cocaine show significant stress-like responses in behavioral tests that follow the cessation of drug administration, and these responses are reversed by administration of a CRF antagonist directly into the brain. For example, injection of a CRF antagonist into the central nucleus of the amygdala reverses the stress-like effects of withdrawal from alcohol, cocaine, nicotine, opioid, and THC in rats. Withdrawal from alcohol, cocaine, nicotine, opioids, and THC is associated with an increase in the release of CRF into the amygdala, suggesting that drug withdrawal can activate CRF systems previously implicated in behavioral responses to stress (Fig. 41.10). Also, stress-induced reinstatement of drug self-administration following extinction can be reversed by CRF antagonists (Shalev et al., 2002).

Norepinephrine in the limbic forebrain also has been associated with the physical and motivational signs of opioid withdrawal. Projections to the bed nucleus of the stria terminalis—a critical part of the extended amygdala—may have a role in the motivational effects of opioid withdrawal (Delfs, Zhu, Druhan, & Aston-Jones, 2000). In addition, the bed nucleus of the stria terminalis is rich in CRF neurons and has been implicated in a feed-forward norepinephrine-CRF-norepinephrine system that may be involved in stress and drug dependence. In this hypothesized feed-forward system, CRF activates brainstem noradrenergic activity, which in turn activates forebrain CRF activity, effectively closing the loop. This mechanism could explain the potentiation of stress responses with repeated exposure that could lead to psychopathology.

Intriguingly, activation of the dynorphin-κ opioid system has long been associated with the development of psychostimulant addiction. Activation of dopamine receptors in the nucleus accumbens shell stimulates a cascade of events that ultimately leads to cyclic adenosine monophosphate response element binding protein (CREB) phosphorylation and subsequent alterations in gene expression, notably the increased expression of protachykinin and prodynorphin mRNA. Subsequent activation of dynorphin systems has been hypothesized to result in decreased dopamine release, thereby contributing to the dysphoric syndrome associated with cocaine dependence (Nestler, 2001). Dynorphins produce aversive dysphoric-like effects in animals and humans and have been hypothesized to mediate negative emotional states by decreasing the functional activity of the mesolimbic dopamine system. However, such enhanced dynorphin action may also activate brain stress responses, such as CRF, or CRF in turn may activate dynorphin (McLaughlin, Marton-Poovici, & Chavkin, 2003; Land et al., 2008). Thus, activation of dynorphin systems by excessive drug use may drive both within-system (decreased reward) and between-system neuroadapatations (increased stress).

Addiction theories such as the opponent process theory postulate that the processes of tolerance to the positive affective state (hedonic tolerance) and subsequent development of the negative affective state (affective withdrawal) play important roles in the transition from drug use to drug dependence (Solomon, 1977). Whereas initial drug use may be motivated by the positive affective state produced by a drug and potentiated by sensitization, continued drug use leads to neuroadaptation to the presence of the drug, the development of a negative affective state, and ultimately self-medication of that affective state. Some theorists have gone so far as to argue that the presence of a negative affective state is the defining feature of addiction and that dysregulation of the residual reward system represents a state of allostasis, the process of obtaining stability through change (Box 41.6). The *allostasis* hypothesis suggests that *both* sensitization and opponent process mechanisms may contribute to changes in incentive motivation (or incentive salience) that convey the vulnerability to relapse for drug addiction.

Neuroadaptation, Prolonged Abstinence, and Relapse

An explanation for the chronic relapsing nature of drug addiction is that reinforcement can be derived from secondary sources that motivate continued drug use. Both positive and negative affective states can become associated with stimuli in the drug-taking environment through classical conditioning processes. When the drug taker is reexposed to these conditioned stimuli, compulsive drug use is reestablished (Box 41.7).

Conditioning to the positive affective states induced by drugs has been demonstrated in animals. Stimuli associated with drugs of abuse can maintain responding in rats and monkeys when presented without the drug. It is also possible to demonstrate conditioned withdrawal. For example, when previously neutral stimuli are paired with naloxone-induced withdrawal, the neutral stimuli themselves elicit signs of opioid withdrawal. Conditioned withdrawal has been observed in opioid-dependent animals and humans.

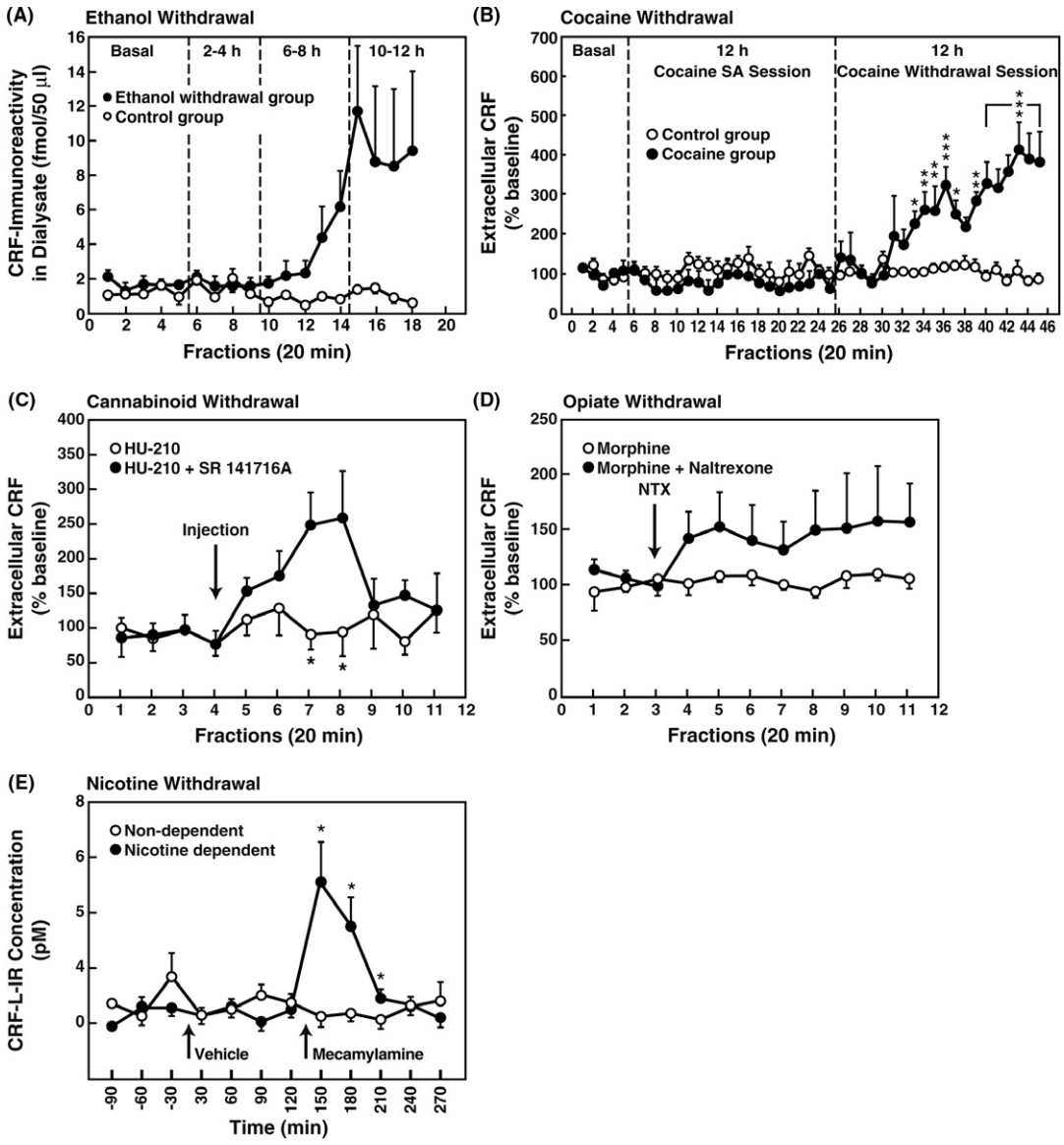


FIGURE 41.10 (A) Effects of ethanol withdrawal on CRF-like immunoreactivity (CRF-L-IR) in the rat amygdala determined by microdialysis. Dialysate was collected over four 2 h periods regularly alternated with nonsampling 2 h periods. The four sampling periods corresponded to the basal collection (before removal of ethanol), and 2–4 h, 6–8 h, and 10–12 h after withdrawal. Fractions were collected every 20 min. Data are expressed as mean ± SEM ($n = 5$ per group). Analysis of variance confirmed significant differences between the two groups over time ($p < 0.05$). *Taken with permission from Merlo-Pich et al. (1995).* (B) Mean (± SEM) dialysate CRF concentrations collected from the central nucleus of the amygdala of rats during baseline, 12 h cocaine self-administration, and a subsequent 12 h withdrawal period (cocaine group, $n = 5$). The control group consisted of rats with the same history of cocaine self-administration training and drug exposure but not given access to cocaine on the test day ($n = 6$). Data are expressed as percentages of basal CRF concentrations. Dialysates were collected over 2 h periods alternating with 1 h nonsampling periods shown by the timeline at the top. During cocaine self-administration, dialysate CRF concentrations in the cocaine group were decreased by about 25% compared with control animals. In contrast, termination of access to cocaine significantly increased CRF efflux that began approximately 5 h post-cocaine and reached about 400% of presession baseline levels at the end of the withdrawal session. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, simple main effects after overall mixed-factorial analysis of variance. *Taken with permission from Richter and Weiss (1999).* (C) Effects of cannabinoid CB_1 receptor antagonist SR 141716A (3 mg/kg) on CRF release from the central nucleus of the amygdala in rats pretreated for 14 days with the CB_1 receptor agonist HU-210 (100 mg/kg). Cannabinoid withdrawal induced by SR 141716A was associated with increased CRF release ($^{*}p < 0.005$, $n = 5$–8). Vehicle injections did not alter CRF efflux ($n = 5$–7). Data were standardized by transforming dialysate CRF concentrations into percentages of baseline values based on averages of the first four fractions. *Taken with permission from Rodriguez de Fonseca, Carrera, Navarao, Koob, & Weiss, (1997).* (D) Effects of morphine withdrawal on extracellular CRF in the central nucleus of the amygdala. Withdrawal was precipitated by administration of naltrexone (0.1 mg/kg) in rats prepared with chronic morphine pellet implants. *Taken with permission from Weiss et al. (2001).* (E) Effect of mecamylamine-precipitated (1.5 mg/kg, i.p.) nicotine withdrawal on extracellular levels of CRF-like immunoreactivity in the central nucleus of the amygdala measured by *in vivo* microdialysis in chronic nicotine pump-treated (nicotine-dependent, $n = 7$) and chronic saline pump-treated (nondependent, $n = 6$) rats. $^{*}p < 0.05$, compared with nondependent. *Taken with permission from George et al. (2007).*

BOX 41.6

ALLOSTASIS

Counteradaptive processes, such as opponent process, that are part of normal homeostatic limitations of reward function can fail to return within the normal range and are hypothesized to contribute to the allostatic state of addiction. Allostasis from the addiction perspective is defined as the process of maintaining apparent reward function stability by changes in brain reward mechanisms. The allostatic state is fueled not only by dysregulation of reward circuits *per se*, but also by the activation of brain and hormonal stress systems (Koob & Le Moal, 2001). The following definitions apply: (1) *allostasis*, the process of achieving stability through change; (2) *allostatic state*, a state of chronic deviation of the regulatory system from its normal, or homeostatic, operating level; and (3) *allostatic load*, the cost to the brain and body of the deviation, accumulating over time, and reflecting in many cases pathological states and accumulation of damage.

BOX 41.7

CRAVING

Craving is a hypothetical construct that can be defined as the memory of the rewarding effects of a drug, superimposed on a negative motivational state (Markou, Kosten, & Koob, 1998). Human studies of craving have been fraught with difficulty in linking it to relapse *per se* in humans with drug addiction, but it is a ubiquitous construct that pervades the fabric of the clinical setting (Tiffany, Carter, & Singleton, 2000). Two types of craving can be conceptualized and linked to animal models.

Craving Type 1: Craving is induced by stimuli that have been paired with drug self-administration such as environmental cues. Termed conditioned positive reinforcement in experimental psychology, an animal model of Craving Type 1 is cue-induced reinstatement where a cue previously paired with access to drug reinstates responding for a lever that has been extinguished. Another example is drug-induced reinstatement where a small dose of the drug elicits responding for a lever that has been extinguished.

Craving Type 2: Craving is conceptualized as a state change characterized by anxiety and dysphoria or a residual negative affective state that combines with Craving Type 1 situations to produce relapse to drug seeking. Animal models of Craving Type 2 include stress-induced reinstatement of drug seeking after extinction or increased drug taking in animals during or after a prolonged deprivation. Craving Type 2 can be conceptualized as a residual negative emotional state or an acute stress-like state, both of which can elicit drug-seeking behavior.

Neurobiological substrates for conditioned drug effects have been investigated in some detail (Cardinal, Parkinson, Hall, & Everitt, 2002; Everitt, Cardinal, Hall, Parkinson, & Robbins, 2000). The enhancement of conditioned reinforcement produced by psychomotor stimulants appears to involve the basolateral amygdala and its glutamatergic projection to the nucleus accumbens with dopaminergic modulation (Robbins et al., 1989). The neurobiological substrates for conditioned withdrawal also may depend on the basolateral amygdala. In contrast, drug-induced reinstatement of drug-seeking appears to be mediated via a prefrontal cortex glutamatergic projection to the nucleus accumbens with dopaminergic modulation (Kalivas & McFarland, 2003). Another powerful source of relapse in human subjects is exposure to a stressor or the state of stress. Both conditions may involve activation of CRF, dynorphin, and noradrenergic systems in the central nucleus of the amygdala and bed nucleus of the stria terminalis. Another powerful source of neurobiological changes that leads to relapse is the loss of frontal cortical function in subregions that control impulsivity and delay of gratification. Damage to such circuits can perpetuate the vulnerability to relapse occasioned by impulsive acts and choices.

Imaging Brain Circuits Involved in Addiction

Brain imaging studies using positron emission tomography with ligands for measuring oxygen utilization or glucose metabolism or using functional magnetic resonance imaging techniques have provided important insights into changes in human brain circuits associated with the development, maintenance, and even vulnerability to addiction. These imaging results, overall, are consistent with circuitry identified in animal studies. Acute intoxication with drugs such as cocaine,

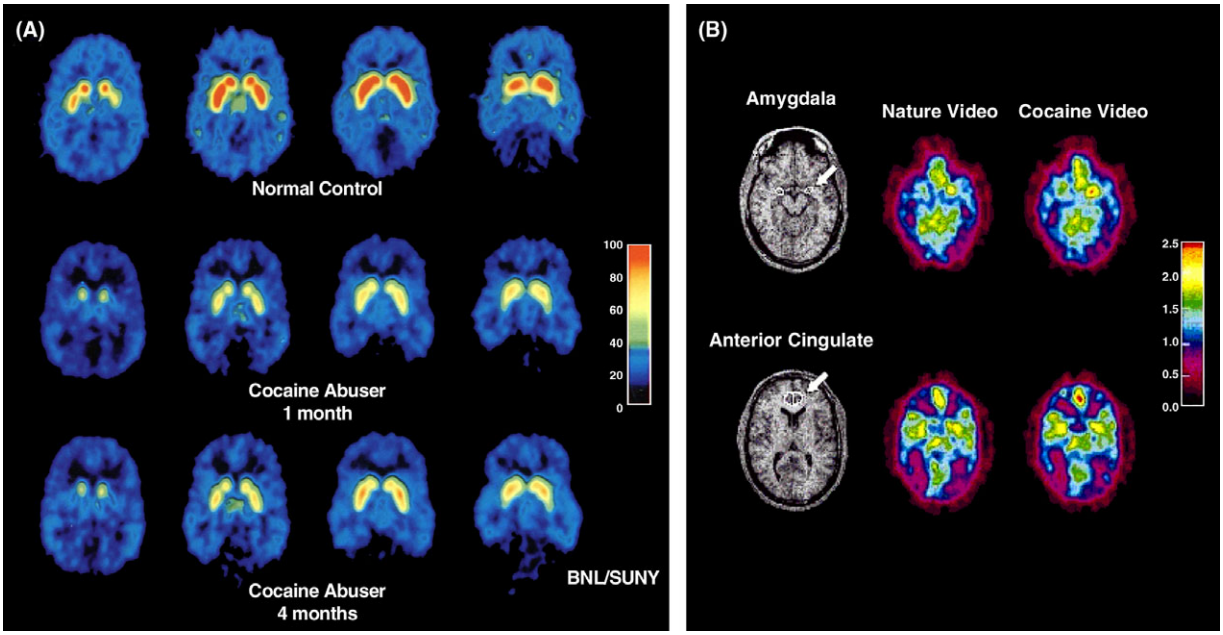


FIGURE 41.11 (A) [^{18}F]*N*-methylspiroperidol images in a normal control and in a cocaine abuser tested 1 month and 4 months after the last cocaine use. The images correspond to the four sequential planes where the basal ganglia are located. The color scale has been normalized to the injected dose. Notice the lower uptake of the tracer in the cocaine abuser compared to the normal control. Notice also the persistence of the decreased uptake even after 4 months of cocaine discontinuation. BNL, Brookhaven National Laboratory; SUNY, State University of New York. *Taken with permission from Volkow et al. (1993).* (B) Transaxial images illustrating the differential increase in relative regional cerebral blood flow (CBF) in the amygdala and anterior cingulate of a detoxified cocaine patient during a nondrug-related (nature) video and a cocaine-related video. Anatomical regions of interest first were localized on the patient's magnetic resonance image (the first image in each row). Region templates were subsequently superimposed on [^{15}O] positron emission tomography (PET) images, yielding radioactive count files for conversion to normalized (relative) regional CBF. The middle and final images in each row show relative regional CBF as measured by PET. The range on the arbitrary color scale is from 0.0 to 2.5 times background activity. Whole brain average regional CBF is 1.0 on the scale. Areas with greatest relative regional CBF are shown in red. Activity in the amygdala and in the anterior cingulate differentially increased during the cocaine video. *Taken with permission from Childress et al. (1999).*

methamphetamine, nicotine, alcohol, and THC produces changes in activation of the orbitofrontal cortex, prefrontal cortex, anterior cingulate cortex, amygdala, and ventral striatum. These changes are reflected both in functional magnetic resonance images and in oxygen utilization as measured by positron emission tomography. The changes evoked during acute intoxication often are accompanied by an increase in the availability of dopamine, particularly in striatal regions.

In contrast, acute and chronic withdrawal reverses these changes, resulting in decreases in metabolic activity in the orbitofrontal cortex, prefrontal cortex, and anterior cingulate cortex and decreases in basal dopamine activity in the striatum and prefrontal cortex as reflected in decreases in D_2 receptor binding (Fig. 41.11). The decreases in dopamine activity have also been validated by challenge studies where cocaine- and alcohol-dependent individuals show blunted dopamine responses to psychostimulant challenge (Volkow et al., 1997b; Martinez et al., 2005).

Craving elicited by presenting cues associated with cocaine, alcohol, nicotine, and THC appears to activate the same circuits as acute intoxication, but with the reactivation of the prefrontal cortex, orbitofrontal cortex, and anterior cingulate cortex and the additional recruitment of activation in the amygdala (Childress et al., 1999) (Fig. 41.11).

Signal Transduction Mechanisms Show Neuroadaptative Changes with Substance Dependence

Chronic drug use not only increases transmitter release via presynaptic effects but also produces long-lasting alterations in postsynaptic receptor activation, signal transduction, and gene expression (Robison & Nestler, 2011). For example, whereas chronic opioid use has minimal effects on the binding of opioid receptors by opioid peptides, it dramatically alters signal transduction in the terminal regions of the mesolimbic dopamine system, such as the nucleus accumbens (Nestler, 2001). This occurs at least in part by increasing adenylate cyclase activity and CREB. CREB is a transcription factor that increases or decreases the transcription of genes. These changes can result in the expression of novel proteins and, consequently, changes in neuronal function that could explain clinical phenomena, such as protracted abstinence and vulnerability to relapse, in people formerly addicted to drugs (Fig. 41.12).

The persistence of changes in drug reinforcement mechanisms that characterize drug addiction suggests that the underlying molecular mechanisms are long-lasting, and considerable attention has been given to the drug regulation of gene expression. Current research focuses on several types of transcription factors, including CREB and novel Fos-like proteins termed chronic Fos-related antigens. FosB, a member of the Fos family of transcription factors that dimerizes with a member of the Jun family to form activator protein-1 transcription factor complexes, has been implicated in the adaptation to chronic exposure to drugs of abuse. Activator protein-1 transcription factor complexes bind to sites in the regulatory regions of many genes. Acute administration of several drugs of abuse activates c-*fos* and Fos-related antigens, but chronic administration of drugs of abuse eliminates these activations and produces a gradual accumulation of FosB. This highly stable protein is hypothesized to function as a sustained molecular switch contributing to vulnerability to relapse after prolonged periods of abstinence (Nestler, 2001). These transcription factors may be possible mediators of chronic drug action. However, it has not yet been possible to relate regulation of a specific transcription factor to specific features of drug reinforcement or addiction.

Genetic and molecular genetic animal models have provided a convergence of data to provide hypotheses regarding vulnerability to addiction in some of the neuropharmacological substrates identified in neurocircuitry studies. High alcohol-preferring rats have been bred that show high voluntary consumption of alcohol, increased anxiety-like responses, and numerous neuropharmacological phenotypes such as decreased dopaminergic activity, increased CRF activity, and decreased neuropeptide Y activity, some of which have support from genetic studies of chromosomal linkage (Carr et al., 1998; McBride, Murphy, Lumeng, & Li, 1990; Murphy et al., 2002).

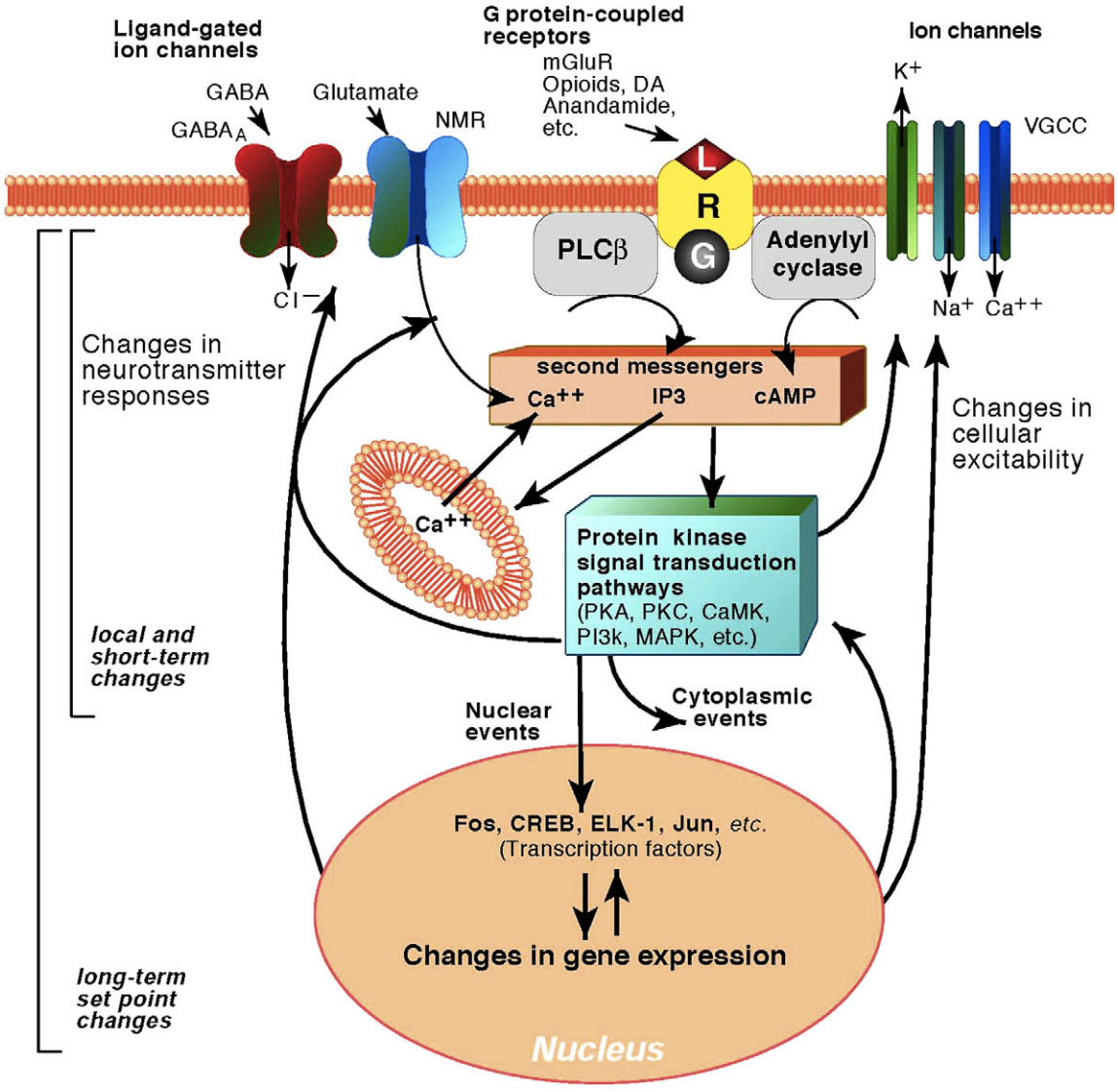


FIGURE 41.12 Molecular mechanisms of neuroadaptation. *Cocaine and amphetamines*, as indirect sympathomimetics, stimulate the release of dopamine which acts at G protein-coupled receptors, specifically D_1, D_2, D_3, D_4, and D_5. These receptors modulate the levels of second-messengers like cyclic adenosine monophosphate (cAMP) and Ca^{2+}, which in turn regulate the activity of protein kinase transducers. Such protein kinases affect the functions of proteins located in the cytoplasm, plasma membrane, and nucleus. Among membrane proteins affected are ligand-gated and voltage-gated ion channels (VGCC). G_i and G_o proteins also can regulate K^+ and Ca^{2+} channels directly through their $\beta\gamma$ subunits. Protein kinase transduction pathways also affect the activities of transcription factors. Some of these factors, like cyclic adenosine monophosphate (cAMP) response element binding protein (CREB), are regulated posttranslationally by phosphorylation; others, like Fos, are regulated transcriptionally; still others, like Jun, are regulated both posttranslationally and/or transcriptionally. While membrane and cytoplasmic changes may be only local (e.g., dendritic domains or synaptic boutons), changes in the activity of transcription factors may result in long-term functional changes. These may include changes in gene expression of proteins involved in signal transduction and/or neurotransmission, resulting in altered neuronal responses. For example, chronic exposure to psychostimulants has been reported to increase levels of protein kinase A (PKA) and adenylyl cyclase in the nucleus accumbens and to decrease levels of $G_{\alpha i}$. Chronic exposure to psychostimulants also alters the expression of transcription factors themselves. CREB expression, for instance, is depressed in the nucleus accumbens by chronic cocaine treatment. Chronic cocaine induces a transition from Fos induction to the induction of the much longer-lasting Fos-related antigens such as ΔFosB. *Opioids*, by acting on neurotransmitter systems, affect the phenotypic and functional properties of neurons through the general mechanisms outlined in the diagram. Shown are examples of ligand-gated ion channels such as the γ-aminobutyric acid-A ($GABA_A$) and glutamate *N*-methyl-D-aspartate (NMDA) receptor (NMR) and G protein-coupled receptors such as opioid, dopamine (DA), or cannabinoid CB_1 receptors, among others. These receptors modulate the levels of second messengers like cAMP and Ca^{2+}, which in turn regulate the activity of protein kinase transducers. Chronic exposure to opioids has been reported to increase levels of PKA and adenylyl cyclase in the nucleus accumbens and to decrease levels of $G_{\alpha i}$. Chronic exposure to opioids also alters the expression of transcription factors themselves. CREB expression, for instance, is depressed in the nucleus accumbens and increased in the locus coeruleus by chronic morphine treatment, while chronic opioid exposure activates Fos-related antigens such as ΔFosB. *Alcohol*, by acting on neurotransmitter systems, affects the phenotypic and functional properties of neurons through the general mechanisms outlined in the diagram. Shown are examples of ligand-gated ion channels such as the $GABA_A$ and the NMDA receptor and G protein-coupled receptors such as opioid, dopamine, or cannabinoid CB_1 receptors, among others. The latter also are activated by endogenous cannabinoids such as anandamide. These receptors modulate the levels of second messengers such as cAMP and Ca^{2+}, which in turn regulate the activity of protein kinase transducers. Such protein kinases affect the functions of proteins located in the cytoplasm, plasma membrane, and nucleus. Among membrane proteins affected are ligand-gated and VGCCs. Alcohol, for instance, has been proposed to affect the $GABA_A$ response via protein kinase C (PKC) phosphorylation. G_i and G_o proteins also can regulate K^+ and Ca^{2+} channels directly through their $\beta\gamma$ subunits. Chronic exposure to alcohol has been reported to increase levels of PKA and adenylyl cyclase in the nucleus accumbens and to decrease levels of $G_{i\alpha}$. Moreover, chronic ethanol induces differential changes in subunit composition in the $GABA_A$ and glutamate inotropic receptors and increases expression of VGCCs. Chronic exposure to alcohol also alters the expression of transcription factors themselves. CREB expression, for instance, is increased in the nucleus accumbens and decreased in the amygdala by chronic alcohol treatment. Chronic alcohol induces a transition from Fos induction to the induction of the longer-lasting Fos-related antigens. *Nicotine* acts directly on ligand-gated ion channels. These receptors modulate the levels of Ca^{2+}, which in turn regulate the activity of protein kinase transducers. Chronic exposure to nicotine has been reported to increase levels of PKA in the nucleus accumbens. Chronic exposure to nicotine also alters the expression of transcription factors themselves. CREB expression, for instance, is depressed in the amygdala and prefrontal cortex and increased in the nucleus accumbens and ventral tegmental area. *Δ9-Tetrahydrocannabinol* (THC), by acting on neurotransmitter systems, affects the phenotypic and functional properties of neurons through the general mechanisms outlined in the diagram. Cannabinoids act on the cannabinoid CB_1 G protein-coupled receptor. The CB_1 receptor also is activated by endogenous cannabinoids such as anandamide. This receptor modulates (inhibits) the levels of second messengers like cAMP and Ca^{2+}, which in turn regulate the activity of protein kinase transducers. Chronic exposure to THC also alters the expression of transcription factors themselves. CaMK, Ca^{2+}/calmodulin-dependent protein kinase; ELK-1, E-26-like protein 1; PLCβ, phosphlipase C β; IP3, inositol triphosphate; MAPK, mitogen-activated protein kinase; PI3K, phosphoinositide 3-kinase; R, receptor. *Modified with permission from Koob, Sanna, and Bloom, (1998).*

BOX 41.8

MAJOR DEPRESSIVE DISORDERS

Feeling sad and blue are common and universal moods experienced in response to transient loss, stress, or disappointment. The clinical diagnosis of major depressive disorder (MDD) involves a persistent symptom complex that includes some combination of depressed mood, diminished interest or pleasure, changes in sleep and appetite, excessive guilt, agitation or retardation, loss of energy, diminished ability to think or concentrate, thoughts of death, and thoughts of suicide. These symptoms are recurrent and persistent and produce impairment.

Major depression is highly prevalent in all cultures and countries studied, and it is the most common psychiatric condition in medical patients. The community lifetime prevalence is approximately 12 to 15%, with the lowest rates in Asian countries and the highest in the Americas, Europe, and Australia. Even in countries with lower rates, it is not rare. The rates are two- to threefold higher in women than men. MDD is uncommon before puberty, with a marked increase in females during adolescence and following childbirth, and a decrease in first onset after menopause. Major depression with a first onset in the 60s or later is often associated with dementia or a medical condition.

Epidemiologic studies from the United States and elsewhere have documented higher lifetime rates of MDD in younger age groups. An analysis by birth cohort showed that rates of depression rose in cohorts born after World War II (the baby-boom generation). Numerous hypotheses noting the social changes that have occurred since that period have

BOX 41.8 (*cont'd*)

been considered, with no confirmed explanation. Follow-up epidemiologic studies carried out a decade or more later, in the 2000s, found the highest prevalence among the middle-age group (30–64 years), suggesting that the cohort born after WWII were carrying forth their high rates as they aged.

Different subtypes of major depression have been identified, including depression with psychotic features and a typical depression. Efforts to understand the heterogeneity of MDD by subtyping symptoms, however, are inherently difficult, as patients' symptoms may change with different episodes. Major depression in the context of a history of mania (bipolar disorder) has the most severe course and requires treatment that prevents both the depressive and the manic symptoms.

Depression is an extremely disabling disorder. About 75% of patients have recurrent episodes. Ten to 30% recover incompletely, with persistent residual symptoms. MDD also complicates the course of cardiovascular illness, diabetes, hypertension, and other chronic medical conditions. Suicide, accidents, and the risk of death from heart disease are highest among the depressed. Because onset is early, recurrence rates are high, and the years lived with disability are high, MDD is considered, according to the World Health Organization, to be among the most disabling medical conditions in the world, similar to ischemic heart disease.

Rates of MDD are highest in persons who are divorced, separated, or widowed, and lowest in unmarried or never-married persons. These differences vary by country and are confounded by the difficulty in determining the sequence of divorce, separation, and depression. The one clear finding is that the ending of marriage, when it occurs, is associated with an increase in MDD.

The increased risk of MDD in first-degree relatives of a depressed patient is approximately two- to sixfold, depending on the severity of the disorder. This pattern of transmission is found across cultures and countries. Early-onset recurrent MDD has the highest relative risk, with estimates as high as eightfold. The best replicated findings are the high rates of MDD, anxiety disorders, and conduct disorder found in the biological young offspring of depressed as compared to nondepressed parents. These studies have been carried out in three generations and show that grandchildren with two generations affected are at extremely high risk of developing MDD at a young age.

The high familial rate of MDD across generations is supported by twin studies showing higher rates of depression in monozygotic twins, and an estimated heritability of 37 to 43%. These estimates are similar to those described for Type 2 diabetes. Several large-scale molecular genetic studies of MDD have been undertaken; most have focused on early-onset (<30 years) recurrent MDD, since this subtype has the highest familial rate. DNA and deidentified clinical information from these studies are stored in the National Institute of Mental Health (NIMH)'s national repositories for use by qualified scientists. As of 2012, a large Genome-Wide Association (GWA) study with participation by numerous investigators across the globe is underway. Many sequence variations in several genes have been reported, but no consistently replicated findings are yet available.

Availability of families at high risk for MDD (by virtue of having a FDR with the disorder) and the numerous positive clinical trials of drugs for MDD have led to a search for biomarkers or "endophenotypes" for MDD. By identifying subjects at risk for depression but not yet ill, this approach allows the disentangling of processes that present *prior* to the illness and are therefore more likely to be true risk factors. One study using magnetic resonance imaging (MRI) found 30% cortical thinning in the second and third generations of high-risk, as compared to low-risk, families that was independent of presence of MDD in the subject. Cortical thinning in areas that regulate emotion and cognitive stress response may thus be a familial trait marker for vulnerability to MDD. The first large-scale study to develop a platform for biomarkers using MRI, electroencephalography, behavioral tasks, and genetics to predict treatment response to MDD using selective serotonin reuptake inhibitors was initiated in 2011. Other promising areas for depression research include studies of circulatory levels of peripheral/serum growth factors and cytokines, which are altered in patients with MDD and which may be normalized or reversed with antidepressant treatment. There is agreement that a single biomarker approach may not be useful. The development of biomarker panels that profile a diverse array of peripheral/serum growth factors, cytokines, hormones, and metabolic markers may provide better coverage of the biologic abnormalities in MDD.

There are numerous studies on the efficacy of both medications and psychotherapy for depression. Treatment can hasten remission, may reduce suicide risk, and, if continued, can prevent relapse. Many of the pharmacologic agents with effective antidepressant action amplify serotonin or neurepinephrine signaling by inhibiting reuptake. There are also several evidence-based psychotherapies developed for MDD. They target the distorted thinking of depression (cognitive therapy) or stressful life events (grief, disputes, transitions) that trigger an episode (interpersonal psychotherapy). Drugs and psychotherapy in combination have been shown usually to have the best outcome, but many patients do not have access to, cannot afford, or do not want both. Both classes of treatment need to deal with the delayed response rate (3 to 4 weeks) and the high recurrence rate. Continuation or maintenance therapy (drugs or psychotherapy) is usually indicated.

Myrna M. Weissman and Ardesheer Talati

Advances in molecular biology have led to the ability to systematically inactivate the genes that control the expression of proteins that make up receptors or neurotransmitter/neuromodulators in the central nervous system using the gene knockout approach. Notable positive results with gene knockout studies in mice have focused on knockout of the μ opioid receptor, which eliminates opioid, nicotine, alcohol, and cannabinoid reward in mice (Contet, Kieffer, & Befort, 2004). Selective deletion of the genes for expression of the dopamine D_1 and D_2 receptor subtypes and the dopamine transporter has revealed significant decreases in activation and reward produced by psychomotor stimulants. Although developmental factors must be taken into account for the compensatory effect of deleting any one or a combination of genes, it is clear that D_1 and D_2 receptors and the dopamine transporter play important roles in the actions of psychomotor stimulants (Caine, Gabriel, Berkowitz, Zhang, & Xu, 2002).

References

American Psychiatric Association, (1994). *Diagnostic and statistical manual of mental disorders,* (4th ed.) Washington, DC: American Psychiatric Press.

Belin, D., Jonkman, S., Dickinson, A., Robbins, T. W., & Everitt, B. J. (2009). Parallel and interactive learning processes within the basal ganglia: Relevance for the understanding of addiction. *Behavioural Brain Research, 199,* 89–102.

Berridge, K. C., & Robinson, T. E. (1998). What is the role of dopamine in reward: Hedonic impact. Reward learning, or incentive salience? *Brain Research Reviews, 28,* 309–369.

Boudreau, A. C., & Wolf, M. E. (2005). Behavioral sensitization to cocaine is associated with increased AMPA receptor surface expression in the nucleus accumbens. *Journal of Neuroscience, 25,* 9144–9151.

Bromberg-Martin, E. S., & Hikosaka, O. (2010). Dopamine in motivational control: Rewarding, aversive, and alerting. *Neuron, 68,* 815–834.

Caine, S. B., Gabriel, K. I., Berkowitz, J. S., Zhang, J., & Xu, M. (2002). Decreased cocaine self-administration in dopamine D-1 receptor knockout mice. *Drug and Alcohol Dependence, 66,* s25.

Caine, S. B., Lintz, R., & Koob, G. F. (1993). Intravenous drug self-administration techniques in animals. In A. Sahgal (Ed.), *Behavioral neuro-science: A practical approach*, (pp. 117–143), Vol. 2. New York, NY: Oxford University Press.

Cardinal, R. N., Parkinson, J. A., Hall, J., & Everitt, B. J. (2002). Emotion and motivation: The role of the amygdala, ventral striatum and prefrontal cortex. *Neuroscience and Biobehavioral Reviews, 26,* 321–352.

Carli, M., Evenden, J. L., & Robbins, T. W. (1985). Depletion of unilateral striatal dopamine impairs initiation of contralateral actions and not sensory attention. *Nature (Lond.), 313,* 679–682.

Carlisle, H. J. (1969). The effects of preoptic and anterior hypothalamic lesions on behavioral thermoregulation in the cold. *Journal of Comparative and Physiological Psychology, 69,* 391–402.

Carr, L. G., Foroud, T., Bice, P., Gobbett, T., Ivashina, J., Edenberg, H., … , Li, T. K. (1998). A quantitative trait locus for alcohol consumption in selectively bred rat lines. *Alcoholism, Clinical and Experimental Research, 22,* 884–887.

Centers for Disease Control and Prevention. (2004). *The health consequences of smoking: A report of the Surgeon General*. Atlanta, GA: US Department of Health and Human Services, CDC.

Childress, A. R., Mozley, P. D., McElgin, W., Fitzgerald, J., Reivich, M., & O'Brien, C. P. (1999). Limbic activation during cue-induced cocaine craving. *American Journal of Psychiatry, 156,* 11–18.

Contet, C., Kieffer, B. L., & Befort, K. (2004). Mu opioid receptor: A gateway to drug addiction. *Current Opinion in Neurobiology, 14,* 370–378.

Delfs, J. M., Zhu, Y., Druhan, J. P., & Aston-Jones, G. (2000). Noradrenaline in the ventral forebrain is critical for opiate withdrawal-induced aversion. *Nature, 403,* 430–434.

Di Chiara, G., & North, R. A. (1992). Neurobiology of opiate abuse. *Trends in Pharmacological Sciences, 13,* 185–193.

Durana, J. H., & Barnes, P. A. (1993). A neurodevelopmental view of impulsivity and its relationship to the superfactors of personality. In W. G. McCown, J. Johnson, & M. B. Shure (Eds.), *The impulsive client: Theory, research, and treatment* (pp. 23–37). Washington, DC: American Psychological Association.

Engel, J., Oreland, L., Ingvar, D. H., Pernow, B., Rossner, S., & Pellborn, L. A. (1987). *Brain reward systems and abuse*. New York, NY: Raven Press.

Everitt, B. J. (1990). Sexual motivation: A neural and behavioral analysis of the mechanisms underlying appetitive copulatory responses of male rats. *Neuroscience and Biobehavioral Reviews, 14,* 217–232.

Everitt, B. J., Cardinal, R. N., Hall, J., Parkinson, J. A., & Robbins, T. W. (2000). Differential involvement of amygdala sub-systems in appetitive conditioning and drug addiction. In J. P. Aggleton (Ed.), *The amygdala: A functional analysis.* (pp. 353–390). Oxford: Oxford Universiry Press.

Everitt, B. J., & Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nature Neuroscience, 8,* 1481–1489.

Fitzgerald, L. W., & Nestler, E. J. (1995). Molecular and cellular adaptations in signal transduction pathways following ethanol exposure. *Clinical Neuroscience, 3,* 165–173.

Flynn, J. P., Vanegas, H., Foote, W., & Edwards, S. (1970). Neural mechanisms involved in a cat's attack on a rat. In R. Whalen, R. F. Thompson, M. Verzeano & N. M. Weinberger (Eds.), *The neural control of behaviour* (pp. 135–173). New York, NY: Academic Press.

George, O., Ghozland, S., Azar, M. R., Cottone, P., Zorrilla, E. P., Parsons, L. H., … , Koob, G. F. (2007). CRF-CRF_1 system activation mediates withdrawal-induced increases in nicotine self-administration in nicotine-dependent rats. *Proceedings of the National Academy of Sciences of the United States of America, 104,* 17198–17203.

Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences of the United States of America, 108,* 15647–15654.

Haber, S. N., Fudge, J. L., & McFarland, N. R. (2000). Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *Journal of Neuroscience, 20,* 2369–2382.

Heimer, L., & Larsson, K. (1966–1967). Impairment of mating behaviour in male rats following lesions in the preoptic-anterior hypothalamic contimuum. *Brain Research, 3,* 248–263.

Hikosaka, O., Sesack, S. R., Lecourtier, L., & Shepard, P. D. (2008). Habenula: Crossroad between the basal ganglia and the limbic system. *Journal of Neuroscience, 28*, 11825–11829.

Hnasko, T. S., Sotak, B. N., & Palmiter, R. D. (2005). Morphine reward in dopamine-deficient mice. *Nature, 438,* 854–857.

Hoebel, B. G. (1974). Brain reward and aversion systems in the control of feeding and sexual behavior. In J. Cole & T. Sonderegger (Eds.), *Nebraska Symposium on Motivation* (pp. 49–112). Lincoln: University of Nebraska Press.

Hyman, S. E., Malenka, R. C., & Nestler, E. J. (2006). Neural mechanisms of addiction: the role of reward-related learning and memory. *Annual Review Neuroscience, 29,* 565–598.

Justinova, Z., Solinas, M., Tanda, G., Redhi, G. H., & Goldberg, S. R. (2005). The endogenous cannabinoid anandamide and its synthetic analog R(+)-methanandamide are intravenously self-administered by squirrel monkeys. *Journal of Neuroscience, 25,* 5645–5650.

Justinova, Z., Tanda, G., Munzar, P., & Goldberg, S. R. (2004). The opioid antagonist naltrexone reduces the reinforcing effects of delta 9 tetrahydrocannabinol (THC) in squirrel monkeys. *Psychopharmacology, 173,* 186–194.

Kalivas, P. W., & McFarland, K. (2003). Brain circuitry and the reinstatement of cocaine-seeking behavior. *Psychopharmacology, 168,* 44–56.

Kelley, A. E., & Berridge, K. C. (2002). The neuroscience of natural rewards: Relevance to addictive drugs. *Journal of Neuroscience, 22,* 3309–3311.

Koob, G. F. (1992). Drugs of abuse: Anatomy, pharmacology, and function of reward pathways. *Trends in Pharmacological Sciences, 13,* 177–184.

Koob, G. F. (1999). Corticotropin-releasing factor, norepinephrine and stress. *Biological Psychiatry, 46,* 1167–1180.

Koob, G. F. (2008). Neurobiology of addiction. In Galanter, M., Kleber, H. D. (Eds.) *Textbook of substance abuse treatment* (4th ed.) (pp. 3–16). Washington, DC: American Psychiatric Publishing.

Koob, G. F. (2011). Theoretical frameworks and mechanistic aspects of alcohol addiction: Alcohol addiction as a reward deficit disorder. In: Spanagel R., & Sommer W. (Eds.) *Behavioral neurobiology of alcohol addiction* (series title: Current Topics in Behavioral Neurosciences). New York, NY: Springer, in press.

Koob, G. F., Bartfai, T., & Roberts, A. J. (2001). The use of molecular genetic approaches in the neuropharmacology of corticotropin-releasing factor. *International Journal of Comparative Psychology, 14,* 90–110.

Koob, G. F., & Bloom, F. E. (1988). Cellular and molecular mechanisms of drug dependence. *Science, 242,* 715–723.

Koob, G. F., Kandel, D., & Volkow, N. D. (2008). Pathophysiology of addiction. In: Tasman, A., Kay, J., Lieberman, J. A., First, M. B., Maj, M. (Eds.) *Psychiatry*, (3rd ed.) (pp. 354–378), Vol. 1. Wiley, Chichester.

Koob, G. F., & Le Moal, M. (1997). Drug abuse: Hedonic homeostatic dysregulation. *Science, 278,* 52–58.

Koob, G. F., & Le Moal, M. (2001). Drug addiction, dysregulation of reward, and allostasis. *Neuropsychopharmacology, 24,* 97–129.

Koob, G. F., & Le Moal, M. (2005). Plasticity of reward neurocircuitry and the 'dark side' of drug addiction. *Nature Neuroscience, 8,* 1442–1444.

Koob, G. F., Riley, S., Smith, S. C., & Robbins, T. W. (1978). Effects of 6-hydroxydopamine lesions to nucleus accumbens septi and olfactory tubercle on food intake, locomotor activity and amphetamine anorexia in the rat. *Journal of Comparative and Physiological Psychology, 92,* 917–927.

Koob, G. F., Sanna, P. P., & Bloom, F. E. (1998). Neuroscience of addiction. *Neuron, 21,* 467–476.

Koob, G. F., & Volkow, N. D. (2010). Neurocircuitry of addiction. *Neuropsychopharmacol Reviews, 35,* 217–238.

Land, B. B., Bruchas, M. R., Lemos, J. C., Xu, M., Melief, E. J., & Chavkin, C. (2008). The dysphoric component of stress is encoded by activation of the dynorphin κ-opioid system. *Journal of Neuroscience, 28,* 407–414.

Leibowitz, S. F. (1980). Neurochemical systems of the hypothalamus: Control of feeding and drinking behaviour and water-electrolyte excretion. In P. J. Morgan & J. Panksepp (Eds.), *Handbook of the hypothalamus* (pp. 299–437), Vol. 3, Part A. New York: NY: Raven Press.

Markou, A., & Koob, G. F. (1992). Construct validity of a self-stimulation threshold paradigm: Effects of reward and performance manipulations. *Physiology and Behavior, 51,* 111–119.

Markou, A., Kosten, T. R., & Koob, G. F. (1998). Neurobiological similarities in depression and drug dependence: A self-medication hypothesis. *Neuropsychopharmacology, 18,* 135–174.

Marshall, J. F., & Teitelbaum, P. (1977). New considerations in the neuropsychology of motivated behaviors. In L. L. Iversen, S. D. Iversen & S. H. Snyder (Eds.), *Handbook of psychopharmacology*, (pp. 201–229), Vol. 7. New York, NY: Plenum.

Martinez, D., Gil, R., Slifstein, M., Hwang, D. R., Huang, Y., & Perez, A. (2005). Alcohol dependence is associated with blunted dopamine transmission in the ventral striatum. *Biological Psychiatry, 58,* 779–786.

Matsumoto, M., & Hikosaka, O. (2009). Two types of dopamine neurons distinctly convey positive and negative signals. *Nature, 459,* 837–431.

McBride, W. J., Murphy, J. M., Lumeng, L., & Li, T. K. (1990). Serotonin, dopamine and GABA involvement in alcohol drinking of selectively bred rats. *Alcohol, 7,* 199–205.

McLaughlin, J. P., Marton-Popovici, M., & Chavkin, C. (2003). κ Opioid receptor antagonism and prodynorphin gene disruption block stress-induced behavioral responses. *Journal of Neuroscience, 23,* 5674–5683.

Melis, M., Spiga, S., & Diana, M. (2005). The dopamine hypothesis of drug addiction: hypodopaminergic state. *International Review of Neurobiology, 63,* 101–154.

Merlo-Pich, E., Lorang, M., Yeganeh, M., Rodriguez de Fonseca, F., Koob, G. F., & Weiss, F. (1995). Increase of extracellular corticotropin-releasing factor-like immunoreactivity levels in the amygdala of awake rats during restraint stress and alcohol withdrawal as measured by microdialysis. *Journal of Neuroscience, 15,* 5439–5447.

Mogenson, G. J., Jones, D. L., & Yim, C. Y. (1980). From motivation to action: Functional interface between the limbic system and the motor system. *Progress in Neurobiology, 14,* 69–97.

Morrison, S. E., & Salzman, C. D. (2010). Re-valuing the amygdala. *Current Opinion in Neurobiology, 20*, 221–230.

Murphy, J. M., Stewart, R. B., Bell, R. L., Badia-Elder, N. E., Carr, L. G., McBride, W. J., Lumeng, L., & Li, T. K. (2002). Phenotypic and genotypic characterization of the Indiana University rat lines selectively bred for high and low alcohol preference. *Behavior Genetics, 32,* 363–388.

National Drug Intelligence Center. (2011). *The Economic impact of illicit drug use on American society [Product No. 2011-Q0317-002]*. Washington DC: U.S. Department of Justice.

Nestler, E. J. (2001). Molecular basis of long-term plasticity underlying addiction. *Nature Reviews Neuroscience, 2,* 119–128.

Olds, J., & Milner, P. (1954). Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *Journal of Comparative and Physiological Psychology, 47,* 419–427.

Parsons, L. H., Koob, G. F., & Weiss, F. (1995). Serotonin dysfunction in the nucleus accumbens of rats during withdrawal after unlimited access to intravenous cocaine. *The Journal of Pharmacology and Experimental Therapeutics, 274,* 1182–1191.

Pfaff, D. W. (1982). *The physiological mechanisms of motivation*. (pp. 217–251). New York, NY: Springer-Verlag.

Pfaus, J. G. (2009). Pathways of sexual desire. *Journal of Sexual Medicine, 6,* 1506–1533.

Phillips, A. G. (1984). Brain reward circuitry: A case for separate systems. *Brain Research Bulletin, 12,* 195–201.

Picciotto, M. R., Zoli, M., Rimondini, R., Lena, C., Marubio, L. M., Pich, E. M., … , Changeux, J. P. (1998). Acetylcholine receptors containing the beta2 subunit are involved in the reinforcing properties of nicotine. *Nature, 391,* 173–177.

Richter, R. M., & Weiss, F. (1999). In vivo CRF release in rat amygdala is increased during cocaine withdrawal in self-administering rats. *Synapse, 32,* 254–261.

Robbins, T. W. (1986). Hunger. In S. L. Lightman & B. J. Everitt (eds.), *Neuroendocrinology* (pp. 252–303). Oxford: Blackwell.

Robbins, T. W., & Everitt, B. J. (1995). Arousal systems and attention. In M. S. Gazzaniga (Ed), *The cognitive neurosciences* (pp. 703–720). Cambridge, MA: MIT Press.

Robbins, T. W., Taylor, J. R., Cador, M., & Everitt, B. J. (1989). Limbic-striatal interactions and reward-related processes. *Neuroscience and Biobehavioral Reviews, 13,* 155–162.

Roberts, A. J., Cole, M., & Koob, G. F. (1996). Intra-amygdala muscimol decreases operant ethanol self-administration in dependent rats. *Alcoholism, Clinical and Experimental Research, 20,* 1289–1298.

Robinson, T. E., & Berridge, K. C. (1993). The neural basis of drug craving: An incentive-sensitization theory of addiction. *Brain Research Reviews, 18,* 247–291.

Robison, A. J., & Nestler, E. J. (2011). Transcriptional and epigenetic mechanisms of addiction. *Nature Review Neuroscience, 12,* 623–637.

Rodd, Z. A., Bell, R. L., Zhang, Y., Murphy, J. M., Goldstein, A., Zaffaroni, A., … , McBride, W. J. (2005). Regional heterogeneity for the intracranial self-administration of ethanol and acetaldehyde within the ventral tegmental area of alcohol-preferring (P) rats: Involvement of dopamine and serotonin. *Neuropsychopharmacology, 30,* 330–338.

Rodriguez de Fonseca, F., Carrera, M. R. A., Navarro, M., Koob, G. F., & Weiss, F. (1997). Activation of corticotropin-releasing factor in the limbic system during cannabinoid withdrawal. *Science, 276,* 2050–2054.

Rolls, E. T. (1985). The neurophysiology of feeding. In M. Sandler & T. Silverstone (Eds.), *Psychopharmacology and food,* (pp. 1–16). Oxford: Oxford University Press.

Sarnyai, Z., Shaham, Y., & Heinrichs, S. C. (2001). The role of corticotropin-releasing factor in drug addiction. *Pharmacological Reviews, 53,* 209–243.

Satinoff, E. (1982). Are there similarities between thermoregulation and sexual behaviour? In D. W. Pfaff (Ed.), *The physiological mechanisms of motivation* (pp. 217–251). New York, NY: Springer-Verlag.

Schultz, W., Romo, R., Ljungberg, T., Mirenowicz, J., Hollerman, J. R., & Dickinson, A. (1995). Reward-related signals carried by dopamine neurons. In J. R. Houk, J. L. Davis & D. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 233–248). Cambridge, MA: MIT Press.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275,* 1593–1599.

Schuster, C. R., & Thompson, T. (1969). Self administration and behavioral dependence on drugs. *Annual Review of Pharmacology and Toxicology, 9,* 483–502.

Shalev, U., Grimm, J. W., & Shaham, Y. (2002). Neurobiology of relapse to heroin and cocaine seeking: A review. *Pharmacological Reviews, 54,* 1–42.

Shaw-Lutchman, T. Z., Barrot, M., Wallace, T., Gilden, L., Zachariou, V., Impey, S., … , Nestler, E. J. (2002). Regional and cellular mapping of cAMP response element-mediated transcription during naltrexone-precipitated morphine withdrawal. *Journal of Neuroscience, 22,* 3663–3672.

Solomon, R. L. (1977). The opponent-process theory of acquired motivation: The affective dynamics of addiction. In J. D. Maser & M. E. P. Seligman (Eds.), *Psychopathology: Experimental models* (pp. 124–145). San Francisco: Freeman.

Solomon, R. L. (1980). The opponent-process theory of acquired motivation: The costs of pleasure and the benefits of pain. *American Psychology, 35,* 691–712.

Solomon, R. L., & Corbit, J. D. (1974). An opponent-process theory of motivation: I. Temporal dynamics of affect. *Psychological Reviews, 81,* 119–145.

Stein, J. J. (2001). *Substance abuse: The nation's number one health problem*. Schneider Institute for Health Care Policy, Robert Wood Johnson Foundation, Princeton NJ.

Substance Abuse and Mental Health Services Administration (2010). Results from the 2009 National Survey on Drug Use and Health: Volume I. Summary of National Findings (Office of Applied Studies, NSDUH Series H-38A, HHS Publication No. SMA 10-4856). Rockville, MD.

Tabakoff, B., & Hoffman, P. L. (1992). Alcohol: Neurobiology. In J. H. Lowenstein, P. Ruiz, & R. B. Millman (Eds.), *Substance Abuse: A Comprehensive Textbook* (2nd Ed), (pp. 152–185). Baltimore, MD: Williams & Wilkins.

Tiffany, S. T., Carter, B. L., & Singleton, E. G. (2000). Challenges in the manipulation, assessment and interpretation of craving relevant variables. *Addiction, 95,* s177–s187.

Ungerstedt, U. (1971a). Adipsia and aphagia after 6-hydroxydopamine induced degeneration of the nigrostriatal dopamine system. *Acta physiologica Scandinavica. Supplementum, 367,* 95–122.

Ungerstedt, U. (1971b). Striatal dopamine release after amphetamine or nerve degeneration revealed by rotational behavior. *Acta physiologica Scandinavica. Supplementum, 367,* 49–68.

Vanderschuren, L. J., & Everitt, B. J. (2004). Drug seeking becomes compulsive after prolonged cocaine self-administration. *Science, 305,* 1017–1019.

Volkow, N. D., Fowler, J. S., Wang, G. J., Hitzemann, R., Logan, J., Schlyer, D. J., … , Wolf, A. P. (1993). Decreased dopamine D2 receptor availability is associated with reduced frontal metabolism in cocaine abusers. *Synapse, 14,* 169–177.

Volkow, N. D., Wang, G. J., Fowler, J. S., Logan, J., Gatley, S. J., & Hitzemann, R. (1997b). Decreased striatal dopaminergic responsiveness in detoxified cocaine-dependent subjects. *Nature, 386,* 830–833.

Walker, B. M., Rasmussen, D. D., Raskind, M. A., & Koob, G. F. (2008). α_1-Noradrenergic receptor antagonism blocks dependence-induced increases in responding for ethanol. *Alcohol, 42,* 91–97.

Walker, S. C., Robbins, T. W., & Roberts, A. C. (2009). Response disengagement on a spatial self-ordered sequencing task: Effects of regionally selective excitotoxic lesions and serotonin depletion within the prefrontal cortex. *Journal of Neuroscience, 29,* 6033–6041.

Weiss, F., Ciccocioppo, R., Parsons, L. H., Katner, S., Liu, X., Zorrilla, E. P., … , Richter, R. R. (2001). Compulsive drug-seeking behavior and relapse: neuroadaptation, stress, and conditioning factors. In V. Quinones-Jenab (Ed.), *The biological basis of cocaine addiction* (series title: Annals of the New York Academy of Sciences). Vol. 937, (pp. 1–26). New York, NY: New York Academy of Sciences.

Weiss, F., Markou, A., Lorang, M. T., & Koob, G. F. (1992). Basal extracellular dopamine levels in the nucleus accumbens are decreased during cocaine withdrawal after unlimited-access self-administration. *Brain Research, 593,* 314–318.

Weiss, F., Parsons, L. H., Schulteis, G., Hyytia, P., Lorang, M. T., Bloom, F. E., & Koob, G. F. (1996). Ethanol self-administration restores withdrawal-associated deficiencies in accumbal dopamine and 5-hydroxytryptamine release in dependent rats. *Journal of Neuroscience, 16,* 3474–3485.

Wise, R. (1982). Neuroleptics and operant behavior: The anhedonia hypothesis. *Behavioral and Brain Sciences, 5,* 39–87.

Wolffgramm, J., & Heyne, A. (1995). From controlled drug intake to loss of control: The irreversible development of drug addiction in the rat. *Behavioural Brain Research, 70,* 77–94.

George F. Koob, Barry J. Everitt, and Trevor W. Robbins

SECTION VII

BEHAVIORAL AND COGNITIVE NEUROSCIENCE

This page intentionally left blank

CHAPTER

42

Human Brain Evolution

HUMAN BRAIN EVOLUTION

Modern evidence concerning brain evolution from comparative studies shows that brain structure, far from being uniform across species, exhibits remarkable variations. This chapter focuses on the evolutionary history of three groups of vertebrates that are of special interest to people: mammals, primates, and humans themselves. The major steps are outlined in the evolution of our large, complex, and extremely useful brains from the smaller, simpler brains of the first mammals, with an emphasis on the neocortex, since this part of the brain is disproportionately large in humans and is critically involved in mental activities and processes that are considered to be distinctly human. We review new evidence about the changes in the internal organization of brains that took place after the human lineage diverged from the lineage leading to our closest relatives, chimpanzees.

THE BRAINS OF EARLY MAMMALS

As early as 340 mya, early reptiles, now most often called "stem amniotes," evolved from amphibians and diverged into two major clades, the sauropsids, which led to present-day reptiles and birds, and the synapsids, which left no survivors other than mammals (Fig. 42.1). We know from the fossil record that early mammals were mainly small, mouse- to cat-sized, and the endocasts of their skulls (Box 42.1) indicate that they had small brains in proportion to their body sizes. In addition, their forebrains were dominated by a large olfactory bulb and a proportionately large expanse of olfactory cortex. In contrast, neocortex was restricted to a small cap on the dorsal surface of the forebrain that failed to extend over the midbrain, as it does in most extant mammals. Thus, olfaction was a very important sense in early mammals, and neocortex was not the dominant structure that it is in most extant mammals. In addition, most early mammals appeared to have been nocturnal with rather poor vision. However, early mammals evolved small bones in the middle ear, the malleus and incus, which together with the stapes transmit oscillations of air pressure to the cochlea of the inner ear. This, and modifications of the cochlea, allowed early mammals to extend their hearing into high frequencies, permitting communication signals that would be undetected by reptilian predators (Allman, 1999). As auditory processing became more important, the inferior colliculus of the auditory midbrain became larger than its visual counterpart, the superior colliculus (Fig. 42.2). As with many extant mammals, touch receptors for facial vibrissae and for the tongue and lips probably provided much of the somatosensory input.

The neocortex of early mammals was perhaps the part of the brain most changed from early ancestors. Neocortex is not, as its name implies, a new part of the brain, as neocortex is homologous to the dorsal cortex of reptiles. Thus, a forebrain structure in early amniotes evolved into neocortex in the sauropsid line leading to mammals and to dorsal cortex in the line leading to modern reptiles and birds (the hyperpallium or wulst in birds). Although it is not known exactly how the dorsal cortex of early amniotes evolved into the neocortex of early mammals, the dorsal cortex of the sauropsid ancestors of mammals is generally thought to resemble the dorsal cortex of present-day reptiles, such as turtles (Fig. 42.3). The dorsal cortex of reptiles consists of a single layer of pyramidal neurons capped by a layer of input axons mixed with dendrites of the pyramidal neurons and a few (10–20 percent) largely inhibitory intrinsic neurons. The fiber layer under the pyramidal cells consists of the axons leaving the pyramidal cells. This type of laminar and cellular organization is also found in medial cortex of reptiles that is homologous to the hippocampal formation of mammals and birds, and the lateral or

© 2013 Elsevier Inc. All rights reserved.

BOX 42.1

HOW DO WE LEARN ABOUT BRAIN EVOLUTION?

There are three main ways to learn about how different brains have evolved. First, the fossil record can be studied. Because bones readily fossilize, whereas soft tissues seldom do, we know a lot about the bones of our ancestors, but much less about everything else. Of course one can infer much about some soft tissues, such as muscles, from their effects on bones, and this is true for brains as well. The brains of mammals fill the skull tightly, and thus the skull cavity of fossils (the endocasts) rather closely reflects the size and shape of the brain, and even the locations of major fissures. Much has been learned about the changes in brain size from the fossil record (Jerison, 2007), and we learn more by considering changes in the proportions of brain parts. For example, early primates already differ from most early mammals by having more neocortex in proportion to the rest of the brain, and more neocortex devoted to the temporal lobe where visual processing occurs. This implies a greater emphasis on functions mediated by neocortex and a greater emphasis especially on the processing of visual information.

We can even learn something about the functional organization of the neocortex from fossils. For instance, subdivisions of the body representation in the primary somatosensory cortex are often marked by fissures in the brain, and fissure patterns revealed by endocasts from fossil skulls have been used to suggest specializations of the somatosensory cortex in extinct mammals (Radinsky, 1976). David Van Essen (2007) has proposed that an important factor in the development of brain fissures is the pattern of connections within the developing brain. According to this theory, densely interconnected regions tend to resist separation during brain growth and form bulges (gyri) that limit the separation distance, whereas poorly interconnected regions are free to fold and form fissures (sulci) that would increase the separation distance. Thus, it is because the hand and face regions of the somatosensory cortex are poorly interconnected that a fissure may develop between the two. In a similar manner, the position of the lunate sulcus has been used to indicate the junction of primary with secondary visual cortex in fossil endocasts of hominid skulls (de Sousa et al., 2010). Thus, the locations of brain fissures in fossil endocasts may tell us something about anatomical connections and functional boundaries in the brains of extinct mammals.

Most of what is known about the evolution of brains is based on comparisons of brain organization of different species of present-day (extant) mammals and other vertebrates. As Darwin recognized, each living species represents the living tip of a largely dead branch of an extremely bushy tree of life. By examining other living tips of the tree, we can infer much about the organizations of the brains (and other body parts) of ancestors that occupied the branching points of this tree. Theories of brain evolution, including those of the evolution of the human brain, depend on reconstructing the probable features of the brains of ever more distant relatives. The comparative method depends on (1) examining the brains of suitable ranges of extant species and (2) determining what features they share, and whether these features are shared because they were inherited from a common ancestor or because they evolved separately. The field of cladistics provides guidelines for making such judgments. The choice of species for comparison depends on the question being asked. For example, to deduce what the brains of early mammals were like, one should examine brains from each of the major branches of the mammalian evolutionary tree (Fig. 42.1). In order to learn about early primates, we need to consider the major branches of primate evolution along with mammals that are thought to be closely related to primates such as tree shrews.

Although the brains of some mammals may have retained more primitive characteristics than others, it is unlikely that the brain of any living mammal fully represents an ancestral condition. Brain features need to be evaluated trait by trait in a comparative context, as any particular feature could be primitive (ancestral) or derived (modified). To reconstruct the course of brain evolution, we need to distinguish ancestral and derived characters. We expect modern species to exhibit a mixture of features that have been retained from a range of ancestors from recent-to-ancient, as many features or traits are thought to evolve independently (called mosaic evolution), although features can be linked and evolve together.

A third source of information about brain evolution is based on understanding the mechanisms and modes of brain development and the constraints they impose on evolution. For example, Finlay and Darlington (1995) have presented evidence that brains change in orderly ways, as they get bigger. In general, larger brains have proportionately more neocortex and less brain stem. Finlay and Darlington suggest this is the case because the late-maturing neocortex of large brains grows longer (late makes great). Another difference between large and small brains is that large brains have more neurons and longer connections. The increase in neurons makes it difficult for each neuron in a large brain to maintain the same proportion of connections with other neurons, as do neurons in a small brain, and to maintain the same transmission times over longer axons. Thus, as larger brains evolve, changes in organization are needed to reduce the commitment to connections, especially connections requiring long, thick axons. A deeper understanding of the genetic, developmental, and structural constraints on brain design could allow us to better postulate how brains are likely to change in organization with changes in brain size.

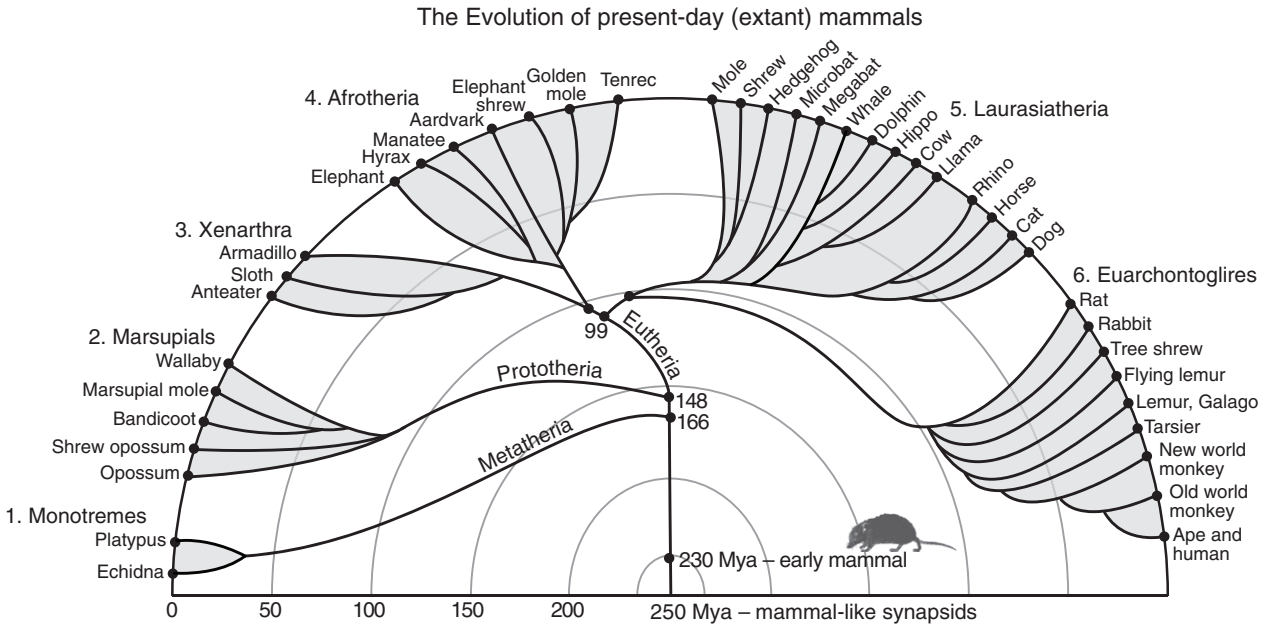


FIGURE 42.1 The phyletic radiation of extant mammals into six major clades or superorders. Most early mammals were shrew-like in appearance. They emerged from mammal-like synapsids about 230 million years ago (mya). Humans and other primates are in the Euarchontoglire superclade. *Based on Murphy, Pevzner, and O'Brien (2004).*

piriform olfactory cortex of mammals, reptiles, and birds (Shepherd, 2011). However, dorsal cortex of mammals evolved new cell layers, to become the familiar six-layered neocortex of mammals. The single layer of pyramidal cells of dorsal cortex was expanded to form the superficial layers 2 and 3, and the deep layers 5 and 6, with each layer specialized for receiving different inputs and providing different outputs, while a middle layer 4 of stellate cells was specialized for receiving fine-grain sensory or other information from the thalamus or other areas of cortex and distributing it to the other layers (Shepherd, 2011). In addition to having the potential to form new kinds of local circuits, the neocortex of early mammals had the potential to be subdivided into any number of functionally different cortical areas that could relate to other areas in various ways, and these cortical areas could be subdivided into sets of repeating modules or columns, providing further computational advantages. As a result, neocortex became a very important part of the brain in some mammals, providing 80 percent of the total brain mass in humans (Azevedo et al., 2009). However, large brains with massive amounts of neocortex came with great costs, including considerable metabolic requirements, and long development times. Larger brains also generate more heat, and are difficult to keep cool (Falk, 1990). In order to have larger amounts of neocortex, a mammal needs to have the potential of a relatively long life and access to high-quality food (Aiello & Wheeler, 1995). Thus, many mammals have small brains and little neocortex with limited computational ability, and likely short lives that are compensated by early maturation, lower metabolic costs, and early reproduction.

Although the neocortex of early mammals was small in comparison to the rest of the brain, it already had several functionally distinct subdivisions. By comparing the organization of neocortex in the brains of members of the six major clades (a clade is any group where all members descended from a common ancestor) of extant mammals (Fig. 42.1), especially those mammals with small brains and relatively little neocortex, we can identify features these brains have in common as a result of sharing an early mammal ancestor. Such a comparative approach leads to several well-supported conclusions (Kaas, 2007a, 2007b, 2011a). First, cortical areas were more similar to each other structurally, and less specialized, than in most extant mammals. While most areas of neocortex had the familiar pattern of six layers of cells, the layers were not markedly different in the types and shapes of intrinsic neurons and pyramidal neurons. Cortical areas were structurally similar, although primary sensory areas were probably already distinguished by having a layer 4 that was more densely packed with small neurons than other cortical areas, and retrosplenial cortex on the medial wall of the cerebral hemisphere was distinguished by having a thin, poorly differentiated layer 4 and a relatively thick layer 2. In general layer 2 was more distinct and layer 1 thicker across cortex than in most extant mammals, reflecting the retention of aspects of dorsal cortex. The structural appearance of neocortex in marsupial opossums (Wong & Kaas, 2009) roughly approximates that expected for early mammals.

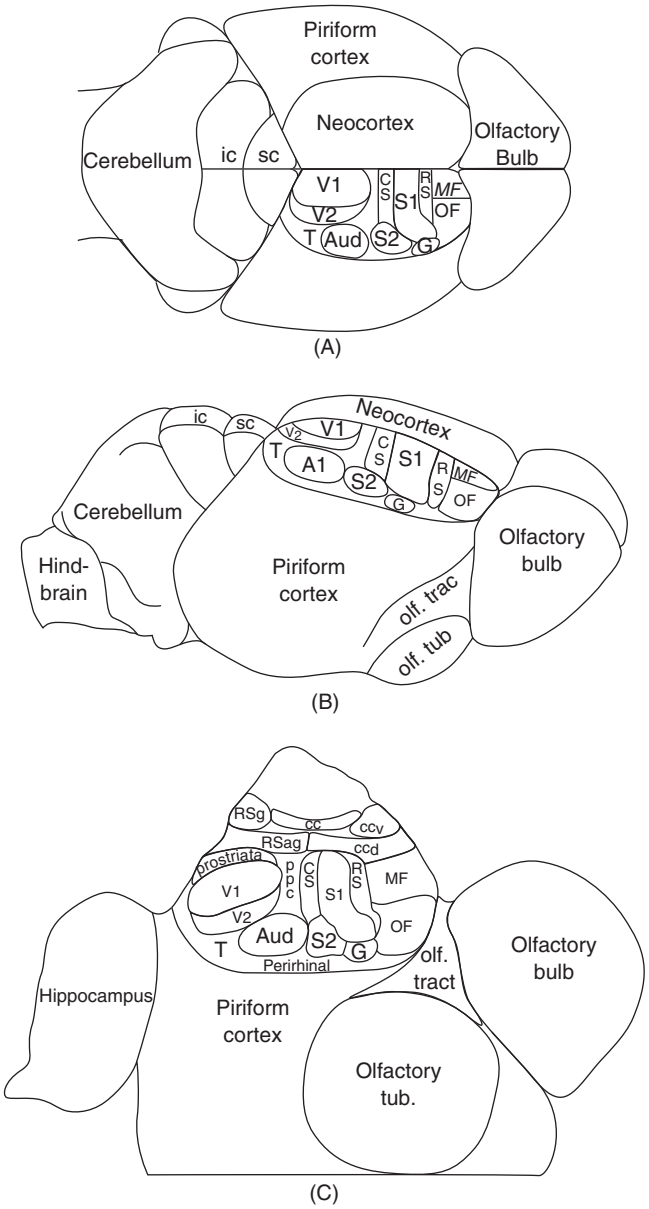


FIGURE 42.2 The inferred organization of neocortex in early mammals. The figure emphasizes the small neocortex in proportion to olfactory piriform cortex and olfactory bulb reflected in the brain endocasts of early mammals. Comparative studies of extant mammals reveal cortical areas that are widespread across mammalian clades, suggesting that they have been retained from early mammalian ancestors. **A.** Dorsal view of the brain with cortical areas indicated. **B**. Lateral view. **C.** View of cortex after it has been separated from the brainstem and unfolded to reveal cortex of the medial wall of the cerebral hemisphere. The hippocampus has been unfolded. Cortical areas include primary and secondary visual areas, V1 and V2, primary and secondary somatosensory areas, S1 and S2, a primary-like auditory core, Aud., caudal and rostral somatosensory belts bordering S1, CS and RS, a gustatory area, G, orbital frontal, OF, and medial frontal, MF, cortex, posterior parietal cortex, PPC, temporal visual cortex, T, dorsal and ventral cingulate cortex, CCd and CCv, and agranular and granular retrosplenial cortex, RSag and RSg, corpus callosum, CC, inferior colliculus, ic, superior colliculus, sc, olfactory tubercle, Olf. Tub., and olfactory tract, Olf. Trac.

We can also conclude that the small cap of neocortex of early mammals was subdivided into a small number of functionally distinct areas, on the order of 10–20, and that these areas have been retained in most lines of descent. From slight regional differences in the laminar appearance of neocortex, early investigators such as Brodmann (1909) surmised that all mammals have functionally significant subdivisions of cortex called areas. Some of these areas are shared by many species (homologous areas retained from a common ancestor), while others are found only in some branches of mammalian evolution. However, it was difficult for Brodmann and other early investigators to reliably identify areas by their differences in histological appearance, especially in the poorly differentiated cortex of most small-brained mammals, but also in the large expanses of rather homogeneous cortex present in large-brained mammals. As a result, areas were often incorrectly delimited within species, and misidentified across species. Fortunately, modern methods of investigation allow many features of cortical biology to be compared in great detail, including patterns of neural connections, ways of representing sensory and other inputs, neuron response properties, and cortical histochemistry. This added information allows cortical areas to be identified within and across species with much more assurance. Nevertheless, many areas remain difficult to identify and delimit with certainty.

When we consider the known distribution of cortical areas in small-brained mammals from the major branches of the mammalian radiation (Fig. 42.1), such as opossums, armadillos, tenrecs, hedgehogs, and even rats (Kaas, 2011a), we find a number of shared areas (Fig. 42.3). Except for a few blind or nearly blind species, such as mole-rats, mammals have a caudally located primary visual area, V1, that is architectonically distinct, has dense inputs from the lateral geniculate nucleus of the visual thalamus, and contains a representation of the contralateral visual hemifield. Other visual areas have been harder to define, but there is widespread evidence across many species for a second visual area, V2, along the lateral border of V1. Thus, it appears that early mammals had a V2, and this V2 has been retained in most present-day mammals. Exceptions are nearly blind mammals, and a few species such as the least shrew that evolved to have extremely small bodies and brains, resulting in the loss of V2 and a few other areas (Catania, Lyon, Mock, & Kaas, 1999). Evidence from comparative studies also supports the conclusion that early mammals also had a visual area along the medial border of V1, a less well-studied area given several names, but commonly called the prostriate area. In addition, one visual area or more was likely located lateral to V2 in temporal cortex of early mammals (T in Fig. 42.2). Because the dorsal cortex of reptiles is predominantly visual in function (Fig. 42.3), one or more visual areas

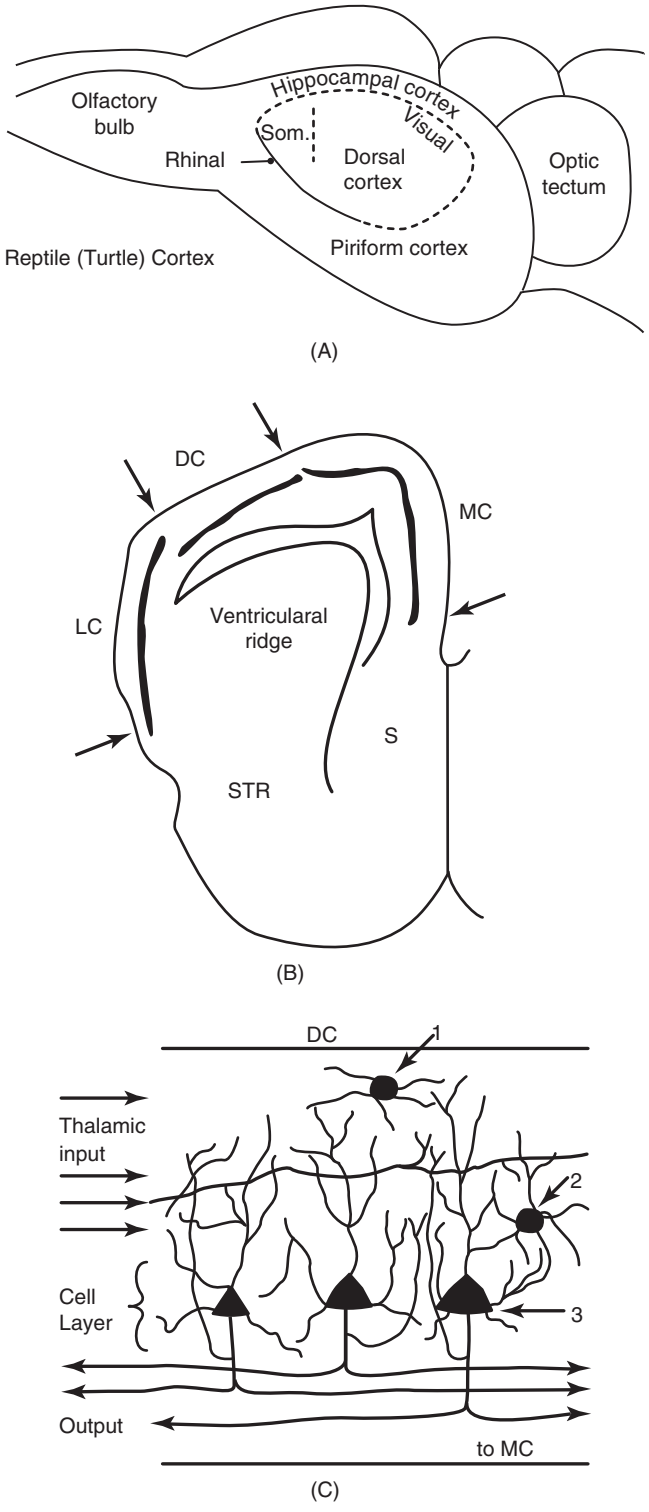


FIGURE 42.3 Dorsal cortex of reptiles. **A.** A dorsolateral view of a turtle brain. The large olfactory bulb provides input to lateral cortex (homologue to piriform cortex of mammals). A slight rhinal dimple is apparent rostrally at the border of dorsal cortex (homologous to neocortex of mammals). Medial or hippocampal cortex (homologous to the hippocampus of mammals) is medial. Most of dorsal cortex receives visual inputs, while a small rostral zone receives somatosensory (som.) inputs. The large optic tectum is part of the midbrain. **B.** A frontal section through the forebrain of a reptile showing the locations of the dorsal cortex (DC), the lateral cortex (LC) and the medial cortex (MC). The solid lines in these cortical areas represent the densely packed pyramidal neurons that form a single cell layer in all three areas. S = septum; STR = striatum. **C.** The cellular structure of dorsal cortex. A densely packed row of pyramidal neurons forms a middle layer. Pyramidal neurons constitute 80 to 90 percent of the neurons in DC. They receive inputs from the visual thalamus as thalamocortical axons stream from lateral to medial in dorsal cortex synapsing on many pyramidal neurons. A few subpial and stellate cells are inhibitory on pyramidal neurons after being activated by pyramidal neuron axons, and input axons. The main branch of pyramidal neuron axon bifurcates with one collateral coursing medially to MC, and the other collateral coursing to the striatum, the visual thalamus, and the optic tectum. Other branches activate other pyramidal neurons and intrinsic neurons. 1. Subpial interneuron. 2. Stellate interneuron. 3. Pyramidal neuron.

likely existed in the dorsal cortex of the synapsid ancestors of mammals.

The organization of auditory cortex has been less studied across a range of mammalian taxa, but a primary or primary-like region has been identified in ventral temporal cortex of all studied mammals (Kaas, 2011b). In many of these mammals, the cortex with primary-like architectonic features includes two or sometimes three different representations of tone frequencies, with one of the representations termed the primary area, A1, and the other, the anterior auditory field among other names. Without further comparative studies, it is uncertain if the primary auditory region of early mammals (Aud) contained a single area, A1, or more. This auditory core area was likely bordered by one or more secondary (belt) auditory fields, but this remains uncertain. Auditory inputs apparently are not represented in dorsal cortex of reptiles and may not have been present in synapsids.

All mammals have several somatosensory areas. A primary area, S1, represents touch on the contralateral half of the body from the tail to the tongue in a mediolateral sequence across anterior parietal cortex. Lateral to S1, a second somatosensory area, S2, has been identified in a wide range of mammals, and the adjoining parietal ventral area, PV, has often been reported, suggesting that S2 and possibly PV (not shown in Fig. 42.2, but located just rostral to S2) were present in early mammals. In addition, narrow bands of cortex along the rostral and caudal borders of S1 with inputs from S1 likely existed, termed here the rostral and caudal somatosensory areas, RS and CS. RS likely had proprioceptive inputs as in area 3a of cats and primates and dysgranular somatosensory cortex of rats. Perhaps surprisingly, early mammals did not have a primary motor area, M1, or any of the premotor areas that characterize primates. This follows from the lack of compelling evidence for an M1 or premotor cortex in marsupials or monotremes, animals that represent early branches of the mammalian radiation. However, inputs from the cerebellum to the anterior thalamus were probably present, and

this motor-related information was distributed to somatosensory cortex, which had motor as well as sensory functions. As dorsal cortex in reptiles appears to have a small region with somatosensory inputs, a region of somatosensory cortex may have been present in the synapsid ancestors of mammals. Finally, there may have been a small region of gustatory cortex, G, located lateral and somewhat rostral to S1 in ancestral mammals, but this is uncertain, as there have been few comparative studies of the representation of taste in the cortex of mammals.

Other cortical areas have been less widely identified in comparative studies, so their presence in early mammals is uncertain. Yet the architectonic features of the granular and agranular retrosplenial areas, RSg and RSa, are distinctive enough to have them be widely identified. There is also reasonably good evidence that early mammals had at least two divisions of cingulate cortex, and two divisions of frontal cortex. Overall, early mammals had a thick neocortex of six layers that was subdivided into as many as 15 or more cortical areas. How this neocortex emerged from the thin, small dorsal cortex with possibly only visual and somatosensory inputs in the ancestors of mammals remains a challenging question (Puelles, 2011).

From this ancestral pattern, a great variety of brain organizations have evolved through alterations in the sizes and numbers of parts, and the connections within and between parts (Kaas, 2007a). Consider, for example, the variations of the primary somatosensory cortex. The duck-billed platypus devotes most of S1 to representing tactile receptors of its highly sensitive bill, and it has added inputs from electroreceptors. The star-nosed mole devotes most of S1 to its long, fleshy nose appendages, rats mostly activate S1 with their facial whiskers, and human S1 has a large representation of the hand, lips, and mouth. In addition, the amount of neocortex varies greatly across species of mammals (Jerison, 2007), and some of this variation is due to differences in numbers of cortical areas present in different groups of mammals (Changizi & Shimojo, 2004; Finlay & Brodsky, 2007; Kaas, 2007a). The number of sensory areas increased independently in several lines of evolution. For example, both cats and monkeys have a large number of visual areas, but the carnivore and primate lines appear to have acquired most of these areas independently rather than from a common ancestor. Thus, many of the visual areas in cats have no homologs in monkeys or humans. A similar situation holds for other regions of the brain. New areas were added, most commonly to sensory and motor regions of cortex rather than to multisensory "association" cortex as once thought. However, sensory areas typically are modulated by other inputs including other sensory and nonsensory inputs. With new cortical areas, new connections between areas and subcortical structures emerged. How new areas evolve remains unclear: they could result from the duplication of existing areas, or by the gradual subdivision of existing areas, or both.

Thus, mammals in various lines of descent evolved new cortical areas and connections, as well as many other specializations of previously existing areas. A long-recognized example of a new brain feature is the corpus callosum, a derived character of placental mammals, having emerged in the ancestors of placentals after they diverged from marsupials. Whereas connections between the two hemispheres are mediated by the anterior commissure in marsupials and monotremes, most of these connections are carried in the shorter, more direct callosal pathway in placental mammals. Primary motor cortex, M1, is another important brain feature that appears to have emerged in early placental mammals after the split from marsupials.

EVOLUTION OF PRIMATE BRAINS

Evolution of Primates

Early primates emerged from small-brained, nocturnal, insect-eating mammals, perhaps as much as 80 million years ago (Martin, 2007), and subsequently branched into two main lineages (Fig. 42.4). One lineage, the strepsirrhines, includes lemurs, lorises, and galagos (collectively, most of the "prosimian" primates). A second lineage, the haplorhines, includes two important daughter branches, one leading to the small, strepsirrhine-like tarsiers, the other leading to the familiar anthropoid primates, which include a New World radiation (the platyrrhines, or New World monkeys) and an Old World radiation (the catarrhines, comprising the Old World monkeys and the ape-human group).

The earliest primates are thought to have been small-bodied, nocturnal visual predators living on insects and small vertebrates, as well as fruit. They adapted to the tropical rainforests by emphasizing vision and visually guided reaching and grasping. This involved having larger, forward-facing eyes, opposable big toes and thumbs, and digits tipped with nails rather than claws. These primates produced the largely nocturnal strepsirrhine radiation with its varied forms, including some species now living in Madagascar that have become diurnal. The haplorhine primates emerged about 60 million years ago in association with a shift from nocturnal to diurnal life, together with an increased emphasis on fruit eating (Ross, 1996). With the shift to diurnal activity came reduced dependence on olfaction, enhancement of the visual system, enlarged body size, and sometimes a more gregarious mode of life. Specifically, the olfactory apparatus was reduced in size, and the eyes enlarged and brought close together. Early haplorhines evolved a fovea, a specialized region of the retina filled with small photoreceptors and devoid of blood vessels that enhances central

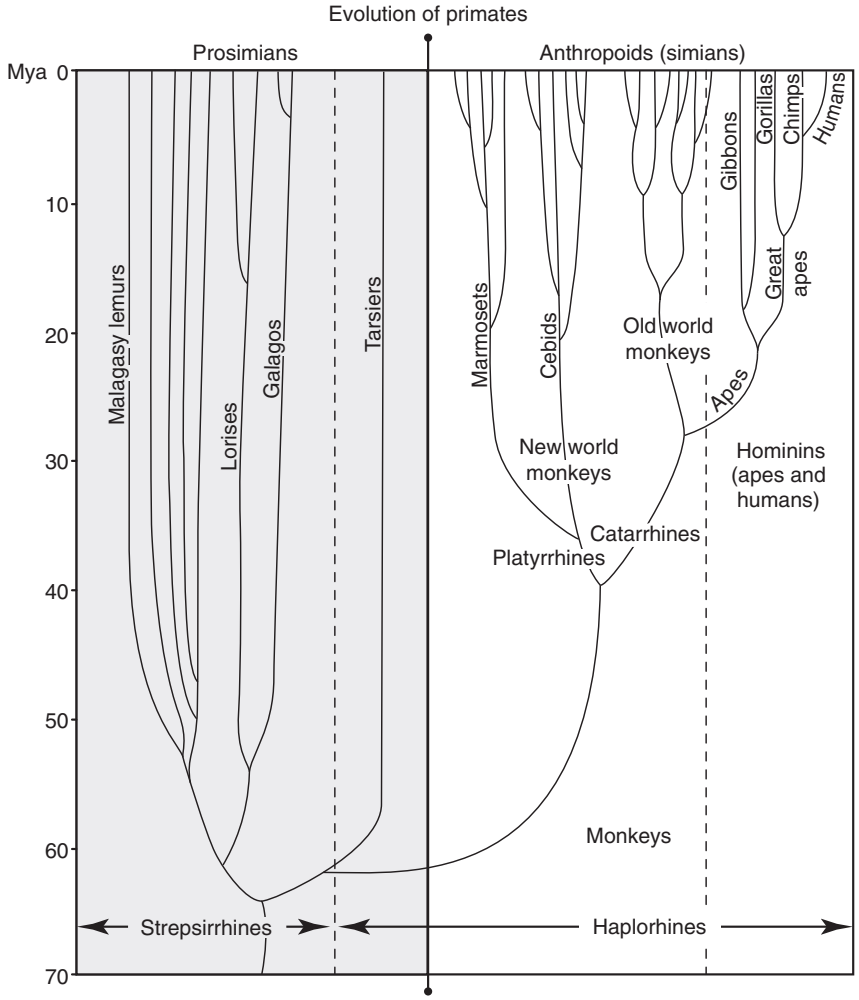


FIGURE 42.4 The evolution and classification of primates. Tarsiers are generally considered to be prosimians, but they are related more closely to anthropoids, so they are recognized as haplorhine primates. Despite the ancient split of strepsirrhines and haplorhines, they share many brain features that are unique to primates. Tree shrews and flying lemurs are thought to be close relatives of primates, and together with them constitute the superorder Archonta. See text.

visual acuity. The reflecting surface at the back of the eye (*tapetum lucidum*), an adaptation to nocturnal vision present in strepsirrhines and many other nocturnal mammals, was lost. Anthropoids underwent further specializations, modifying cone morphology, increasing the proportion of cones to rods, and filling the fovea with cones to the near exclusion of rods. These adaptations enabled anthropoids to achieve a high degree of color discrimination. Nevertheless, whereas humans and other Old World anthropoid primates have three types of retinal cones, ancestral anthropoids probably possessed two types of cones, similar to most modern strepsirrhines and New World monkeys. A third cone type, enabling full trichromatic vision, appears to have evolved independently in the ancestors of Old World anthropoids and in several New World monkey groups. The enhancements of color vision in the anthropoids are plausibly adaptations for distinguishing ripe, edible fruits. The shift to diurnal activity was also marked by larger social groupings, which may offer enhanced protection from predation. These changes were accompanied by increased brain size, including increases in the temporal lobe visual region, and in regions of the parietal and frontal cortex mediating motor control and social interactions.

The ancestors of present-day tarsiers evidently abandoned the diurnal niche to become nocturnal visual predators again. Tarsiers retain a fovea, but they have a rod-dominated retina and their enormous eyes are sensitive at low light levels without the aid of a reflecting tapetum. The primary visual cortex became very large relative to the rest of the brain and is extremely well differentiated, with a multiplicity of layers and sublayers (Wong, Collins, & Kass, 2010). Tarsiers became such extremely specialized visual predators that they eat no plant food. Other haplorhines (anthropoids) remained diurnal and spread to many niches, including those outside the rainforest and niches

based more on eating leaves as well as fruits. Some increased considerably in body size. Later anthropoids were able to process hard, husked fruits with their hands and teeth. Some early anthropoids managed to reach South America from Africa, apparently by rafting, to form the New World monkey radiation. All modern anthropoids are diurnal with the exception of owl monkeys, a New World monkey group that has (like tarsiers) become secondarily nocturnal.

In Africa, early anthropoids diverged some 25 to 30 million years ago into lines leading to modern Old World monkeys and to apes. At first, apes were the most successful radiation, coming to occupy a range of rainforest and open woodland environments, while monkeys were quite rare. This condition changed radically some 10 million years ago when monkeys became abundant and apes rare. This change may have resulted from the advent of a more seasonably variable and drier climate. Ancestral Old World monkeys acquired specialized teeth suited to a leaf and fruit diet in drier climates. Also, the more rapid reproduction of monkeys may have made them more resistant to extinction than apes.

Some 6 to 8 million years ago, a line of apes diverged into two branches: one that gave rise to modern common chimpanzees and bonobos (pygmy chimpanzees) and a second branch that led to the group of bipedal apes that includes modern humans, the "hominins" (formerly known as "hominids," a term that now is used to refer to the great ape-human group) (Fig. 42.5).

Although the hominins today are represented by only one species, *Homo sapiens*, they were a remarkably successful and diverse group (de Sousa & Wood, 2007; Wood & Lonergan, 2008). The oldest known hominin radiation, australopiths, dates back at least 5 million years. These early hominins were bipedal, but skeletal traits suggest they retained considerable ability for climbing trees. Body size was rather small, and males were much bigger than females. Their brains were only slightly larger than those

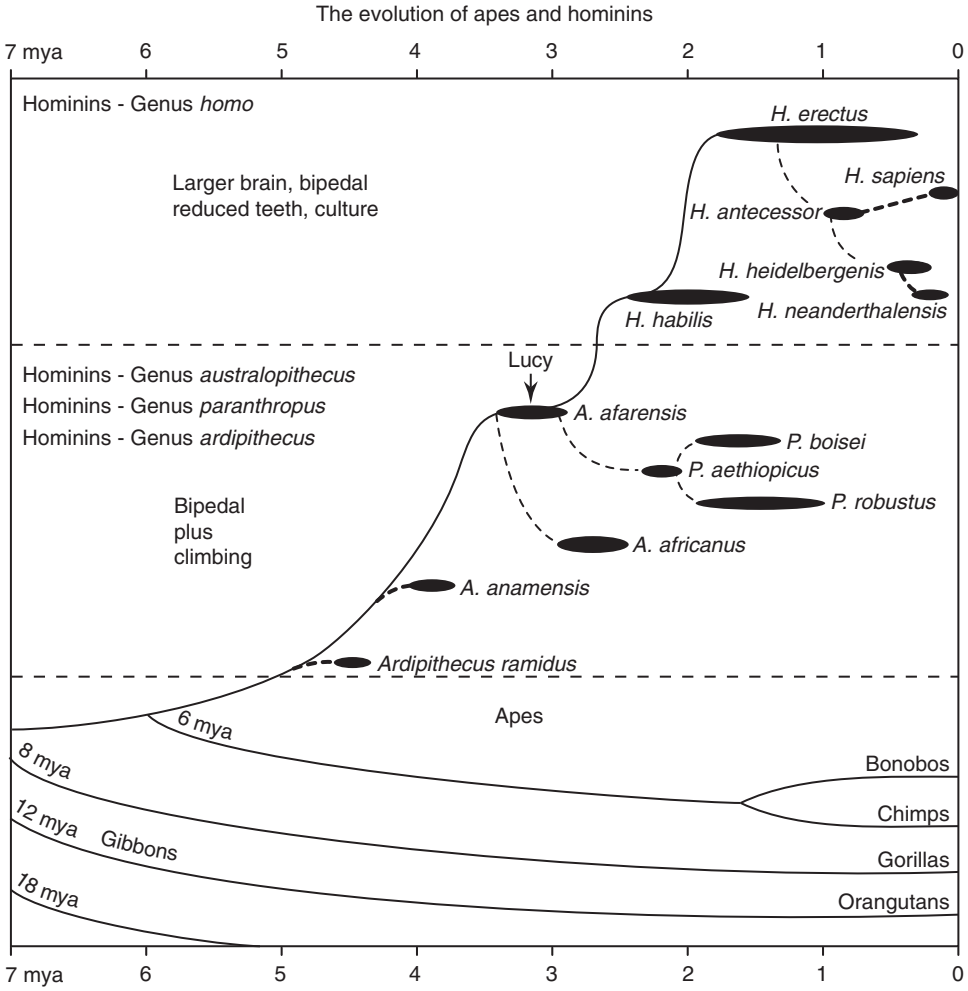


FIGURE 42.5 The evolution of apes and humans. Hominins include those extinct species that are judged to have been bipedal. *Ardipithecus ramidus* is generally thought to be bipedal while *Australopithecus* and *Paranthropus* clearly appear to have been bipedal. The age of the famous fossil, Lucy, is indicated. Relationships are somewhat uncertain, and more branches on the tree exist (Wood & Lonergan, 2008).

of apes of similar body size. Hominin traits may have emerged as adaptations to a drier environment with grassland and savanna. Early australopithecines soon gave rise to a number of species differing in body size and limb proportions, as well as in characteristics of the jaws and teeth and brain size. Our own genus emerged some 2 million years ago with *Homo habilis* (or "handy man," due to its use of stone tools). *Homo habilis* had a slightly increased cranial capacity compared to australopithecines, reduced face and teeth, and pelvic modifications for improved bipedal locomotion and the birth of neonates with larger heads. About 1.7 million years ago, *H. habilis* appears to have been replaced by *H. erectus*, a larger hominin with a further reduction in face and teeth and a larger brain. Shortly thereafter, *H. erectus* spread out of Africa to central Asia. *Homo sapiens* emerged from an African population of *H. erectus* some 250,000 to 300,000 years ago. Other members of the genus *Homo* coexisted with *H. sapiens* until recently. The Neanderthals, who adapted to a colder Europe and southwest Asia over 130,000 years ago, disappeared and were replaced by modern *H. sapiens* some 35,000 years ago. Neanderthals were shorter and more heavily built than modern humans, but had a comparable or slightly larger cranial capacity. As another known example, the fossilized bones of a probable descendant of *H. erectus*, *H. florensensis*, have been dated to 18,000 years old (Morwood et al., 2004). *Homo florensensis* were small (one meter tall and 20 Kg.) with brains (380 grams) smaller than those of the great apes. Over the last 15,000 years, some populations of modern humans have become smaller and have reduced their brain size, possibly as a result of a poorer diet as agriculture replaced hunting and gathering.

Chimpanzees and bonobos (also misleadingly known as "pygmy chimpanzees") appear to be our closest surviving relatives. Humans and chimpanzees are 98 to 99% similar in the coding sequences of genes, although the overall difference in DNA is closer to 96%, the difference reflecting the independent duplication, insertion, and deletion of DNA segments in the two lineages, and changes in noncoding regions that control gene expression. At present, we do not understand what changes in DNA account for the differences in brain size, anatomical organization, and function between humans and other apes, but this is an active area of research.

Presently, primates constitute a highly diversified taxonomic group (clade) that includes 14 families and over 350 species. Describing their brains is a challenging task, as the brains of only a few primate species have been extensively studied. It is clear, however, that primate brains vary enormously in size, and much of that variation reflects the 5000-fold range of body sizes found in primates—from the 40-gram mouse lemur to the 200-kilogram male gorilla (Fig. 42.6). Nevertheless, there are important differences in brain size that cannot be attributed to differences in body

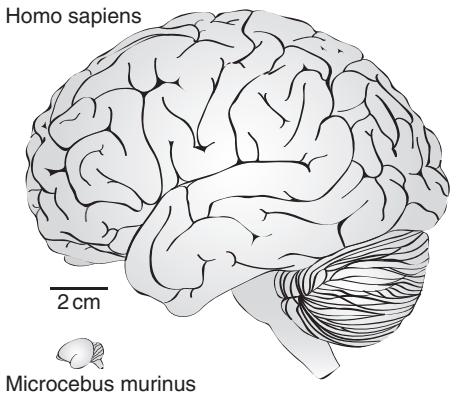


FIGURE 42.6 Lateral views of the brains of a human and a small prosimian primate, the mouse lemur, to illustrate the great range of sizes for present-day primates.

size. These variations in brain size reflect changes in the number of cortical areas and the size of individual areas. In addition, however, the internal architecture and intrinsic connectivity of areas were modified by evolution, as were the ways in which cortical areas are interconnected, reflecting the ways in which primates have diversified behaviorally. Despite these variations, primate brains also share a large number of features, some of which have been retained from nonprimate ancestors, and some of which emerged in early primates and have been widely retained.

Brains of Early Primates

Early primates resembled present-day lemurs and galagos in body form and brain shape, but the brains were significantly smaller in proportion to the body. Except for a lateral fissure and a calcarine fissure, features of the brains of most primates, there were only a few shallow fissures. We know most about the organization of neocortex and other parts of the brain of early primates from studies of present-day strepsirrhine primates, especially galagos, as they have been most extensively studied. The brains are smaller in proportion to body size than the brains of anthropoid primates, and they contain a number of distinctive features that are also found in the brains of other primates. While our focus is on cortical organization, we will also briefly describe subcortical features that were likely present in early primates, especially those for sensory systems.

Vision became a more important sense in early primates (Kaas, 2007b) as they came to rely less on olfaction and more on detecting prey and other food visually, and using eye-to-hand coordination to navigate in the fine branches of bushes and trees and bring food objects to the mouth. Compared to the nearest relatives of primates (tree shrews, flying lemurs, rabbits, and rodents), early primates had more forward-facing eyes, creating a larger binocular field of 130 degrees or more and a shift of the

BOX 42.2

WHY BRAIN SIZE IS IMPORTANT

Larger brains are generally thought to be computationally better because they usually have more neurons. However, growing bigger brains with more neurons creates a need for modifications in brain organization, and some solutions are likely to be common across taxa, allowing predictions about brain organization with brain size (Kaas, 2000). Enlarging a brain by having more neurons creates two related problems. If brain functions depend on each neuron maintaining the same proportion of connections with other neurons, then more neurons means more connections, which would be longer and take up more space. In addition, the conduction times of longer axons and dendrites would be longer, unless the thickness of the axons and dendrites were increased with the added length (Bekkers & Stevens, 1970). Thus, larger brains would have larger neurons with longer, thicker axons and dendrites. This suggests that brains would reach a maximum size, perhaps not much larger than that of present-day humans, because brain connections and glia would take up so much space with the addition of more neurons that such additions would be impractical. However, a design solution for such connection problems is for brains to become increasingly modular with larger size, so that short distance connections predominate, and only a few neurons have the long connections that are needed to coordinate the modular parts. Rodents do seem to increase the average neuron size with brain size, so that neuron densities greatly decrease as proportionately more space is devoted to glia and connections, but primates maintain neuron densities with increasing brain size, implying that average neuron size does not increase (Herculano-Houzel, Collins, Wong, & Kass, 2007). No other mammal, including elephants and whales with much larger brains, has the same number of neurons as the 1.5-kg human brain. A rodent with 86 billion neurons would need a 35-kg brain, well beyond the largest known brain mass of 9 kg for the blue whale (Herculano-Houzel, 2009). Whales and elephants may have evolved more modular brains as they acquired large brains, but they do not seem to have done it to the extent that seems to have occurred in the evolution of the large human brain. Over the course of the evolution of the large human brain, the neocortex of primates, which makes up the bulk of the human brain, acquired more cortical areas and more modular subdivisions of areas as brain sizes increased. As an example, the hemispheric specializations of the human brain presumably reduce the need for long connections between the two hemispheres.

A related issue is that large cortical areas are unlikely to function in the same manner as small cortical areas. Unless neurons compensate with larger dendrites and longer intrinsic connections as areas get larger, the computational window or scope of neurons will decrease (Fig. 42.10). For example, as a visual area gets bigger, individual neurons would evaluate information from less and less of the total visual field. This implies that as areas get bigger, their neurons become less capable of global center-surround comparisons and more devoted to local center-surround comparisons. Thus, some of the integrative functions of large areas must be displaced to smaller areas. The large V1 of the human brain preserves the detail of visual scenes, but perception depends on other, smaller visual areas. It is also apparent that changes in the sizes of dendritic arbors and the lengths of intrinsic axons in smaller areas would have more impact on the sizes of computational windows of neurons. Comparable changes in dendrites and axons would enlarge or reduce receptive field sizes more in a small than a large visual area. Because their functions are more modifiable by smaller structural modifications, smaller areas may be specialized more easily for different functions. Neurons in large areas typically do not have larger dendritic arbors and longer intrinsic connections, and indeed, they may have smaller dendritic arbors. Thus, V1 of primates has overall the smallest and most densely packed neurons of all cortical areas, and this is especially the case in large-brained anthropoid primates (Collins, Airey, Young, Leitch, & Kaas, 2010). In addition, although primary sensory areas are often larger in larger brains, they are not enlarged in proportion to the rest of cortex. Thus, in surface area primary visual cortex is less than three times larger in human brains than in the brains of macaque monkeys, whereas the neocortex as a whole is over 10 times larger. Neocortex in chimpanzee brains is one-fourth the size of the human neocortex, while primary visual cortex (V1) is about the same size. The general lack of proportional growth of cortical areas with brain size reduces the impact of the changing functions of areas with size, and reflects the addition of other smaller cortical areas.

retinal region of high receptor and ganglion cell density toward the center of the retina. These modifications permitted a fine-grain level of stereopsis in central vision and the detection of otherwise cryptic invertebrate prey. The retina had the three major classes of ganglion cells that are found in all extant primates; the small koniocellular (K), medium parvocellular (P), and large magnocellular (M) ganglion cell classes, which are named after their layers of termination in the lateral geniculate nucleus of the visual thalamus. Detailed vision depends on the

P cells, which constitute the majority of retinal ganglion cells in all primates. K cells constitute only about ten percent of the ganglion cells in diurnal primates, but they are more common in nocturnal primates, and would have been so in early primates since they were nocturnal. M cells respond to visual change such as that commonly produced by object movement. M cells represent roughly ten percent of the ganglion cells. The ganglion cells project to a number of small brainstem targets, as well as to the dorsal lateral geniculate nucleus (LGN) and the superior colliculus (SC).

Early primates and all subsequent primates differed from their nonprimate ancestors by representing only the contralateral visual hemifield via the nasal retina of the contralateral eye and the temporal retina of the ipsilateral eye in the superior colliculus. Nonprimate mammals preserve the ancestral mammalian condition, in which the complete retina of the contralateral eye projects to SC. This modification in primates likely promoted the coordination of eye movements and binocular vision.

The other major target of the retina, the LGN, also became specialized with the first primates. A common pattern of four main layers emerged with two ventral magnocellular layers and two dorsal parvocellular layers, with the outer layers for the contralateral eye and the inner layers for the ipsilateral eye. As early primates were nocturnal, they also likely had two thin koniocellular layers between the P layers, one for each eye, as in all extant strepsirrhines. These layers did not persist in anthropoid primates, in which the K cells remained but were scattered between the other layers. In many anthropoid primates, including humans, the P-cell layers became massive, and subdivided into four or more sublayers that allowed further emphasis on detailed object vision. The visual pulvinar was also organized in a distinctive primate specific pattern, with an inferior pulvinar of at least three divisions and a lateral pulvinar of possibly two divisions.

In early primates, the fossil record indicates that visual cortex occupied a large portion of neocortex as occipital cortex extended caudally to cover the midbrain and much of the cerebellum, and temporal cortex extended forward and down as in present-day galagos (Fig. 42.7). This visual cortex was divided into a number of visual areas, some of which appear to be unique to primates. Other visual areas retained from nonprimate ancestors became modified to subserve the increased role for vision in early primates.

The arrangement of some of the visual areas that were likely present in the visual cortex of early primates are shown on a surface view of the neocortex of one hemisphere of a galago brain, after the neocortex has been separated from the rest of the brain and flattened so that hidden surfaces are visible (Fig. 42.7). As in extant primates, early primates had three early stages of cortical processing: V1, V2, and V3. V1 was large and densely packed with small layer 4 cells, more so than any other cortical area (Collins et al., 2010), as a way to promote detailed vision. In primates, but not in the close relatives of primates, V1 has a distribution of cytochrome oxidase "blobs" that may represent a specialized output of the K-geniculate layers (Casagrande & Kaas, 1994). All primates also have an orderly arrangement of direction-of-motion and stimulus-orientation neurons in V1 that look like pinwheels from the surface. This feature of V1 likely predated the evolution of primates, as the close relative of primates, tree shrews, also have orientation-specific pinwheels in V1, but V1 of more distantly related rodents and lagomorphs does not. The segregation of eye of origin information into ocular dominance bands or patches is also a specialization in V1 of primates, being absent in tree shrews, rodents, and rabbits. Ocular dominance columns evolved independently in carnivores, as did blobs.

V2 is another area that is common to most mammals, while being specialized in primates by having three repeating classes of band-like modules that cross the width of V2, the so-called thin, pale, and thick stripes, each receiving a different type of input from V1 and projecting differently to higher order visual areas. These modules exist in strepsirrhine as well as in anthropoid primates, but they are much more histologically distinct in anthropoids. These band-like modules are not found in the relatives of primates, suggesting that they evolved in early primates and were further differentiated in anthropoids.

V3 also exists in all studied primates, although its existence as a field bordering V2, with representations of both the upper and lower visual quadrants, has been questioned. There is no compelling evidence yet for a V3 in the close relatives of primates, despite the clear evidence for a V3 type of area in carnivores, so V3 may be a primate specialization. However, V3 has only recently been well characterized in primates, and the close relatives of primates should be further investigated.

The middle temporal visual area, MT, is a visual area that is recognized widely as being present in all primates. MT receives most of its activation inputs from the M ganglion cell population via direct and indirect projections from V1, and is highly involved in analyzing visual motion. MT is part of a constellation of functionally related areas, MST, FST and MTc (V4t), that all appear to be unique to primates, although homologues of MT in other mammals have been suggested. The dorsolateral visual area (DL or V4) is another area that may be a primate specialization, as is the dorsomedial visual area (DM or V3A). A number of other visual areas have been proposed, especially for macaque monkeys and humans. But for the most part, these areas have not been well defined across a range of primate species. Nevertheless, primates share as many as 10 to 15 visual areas. All primates appear to have their visual systems organized into dorsal and ventral streams of visual processing for guiding actions and object vision, respectively. Dorsal-stream

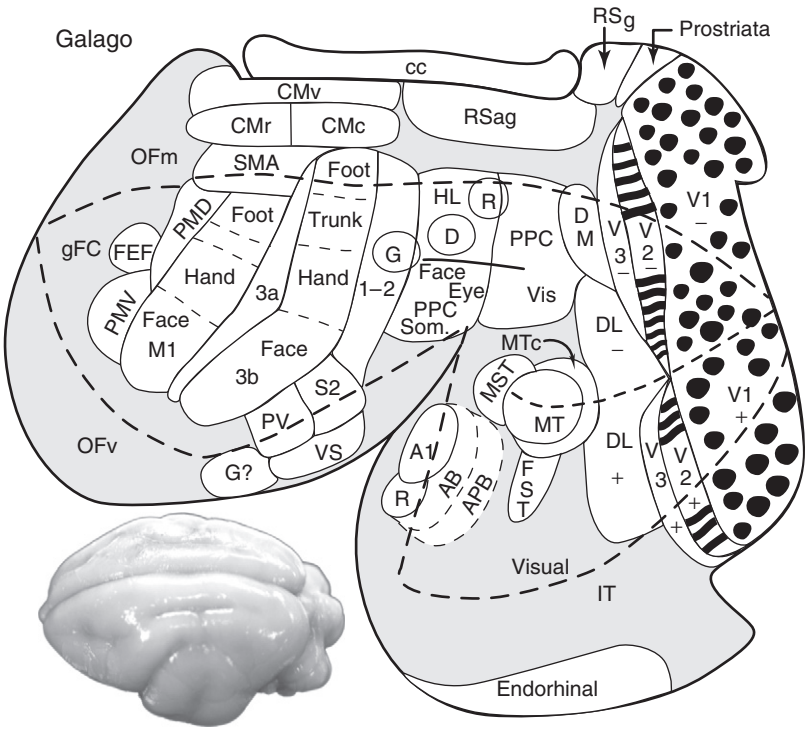


FIGURE 42.7 A surface view of the flattened neocortex of a prosimian primate, *Galago garnetti*, showing some of the proposed visual, somatosensory, auditory, and motor areas. Visual areas include the primary (V1) and secondary (V2) areas, common to most mammals, but with the modular subdivisions (blobs in V1; bands in V2) characteristic of primates. As in other primates, galagos have a third visual area (V3), a dorsomedial area (DM), a middle temporal area (MT), a dorsolateral area (DL), a fundal-sucal-temporal area (FST), a middle superior temporal area (MST), and an inferior temporal visual region (IT) with subdivisions. Posterior parietal cortex (PP) contains a caudal division with visual inputs and a rostral division with sensorimotor functions. The auditory cortex includes a primary field (A1), a rostral area (R), and an auditory belt (AB), which includes several areas, and parabelt auditory cortex (APB). The somatosensory cortex includes a primary area (S1 or 3b), a parietal ventral area (PV), a secondary area (S2), a parietal ventral area (PV), a somatosensory rostral belt (RS or 3a), and a somatosensory caudal belt (CS or possibly area 1 or areas 1 plus 2). A gustatory area (G?) may have been present. Motor areas include a primary area (M1), ventral (PMV) and dorsal (PMD) premotor areas, a supplementary motor area (SMA), a frontal eye field (FEF), and other motor areas on the medial wall of the cerebral hemisphere. The thick dashed line encloses cortex that can be seen on a dorsolateral view of the intact brain (lower left). Granular frontal cortex (gFC); orbital frontal medial cortex (OFm); orbital frontal ventral cortex (OFv); ventral, rostral, and caudal cingulate motor areas (CM); retrosplenial granular and agranular areas (RSg and RSug); corpus callosum (CC). *Modified from Kaas (2007b).*

motion-sensitive visual areas such as MT project to posterior parietal cortex and object related visual areas such as DL/V4 project to visual areas of the inferior temporal lobe. Both parietal and temporal areas relay to frontal cortex (Preuss & Goldman-Rakic, 1991).

The auditory system has been less extensively studied than the visual system in primates, and subcortical processing stations such as the inferior colliculus and medial geniculate complex are assumed to be organized in a manner similar to those of other mammals. However, mammals vary considerably in the arrangement of auditory cortical areas (Kaas, 2011b). Primates are generally thought to have three early stages of auditory processing, a primary core, a surrounding belt, and an adjoining parabelt (Kaas & Hackett, 2000). The primary core in all primates appears to include at least two areas: A1 and a rostral area, R. Both represent tone frequencies from high-to-low, and have the histological, connectional, and functional properties of primary sensory cortex. Little is known about how belt and parabelt regions are organized across primate taxa, or in any strepsirrhine primates, so little can be inferred about early primates. However, given the presence of at least two primary areas in rodents, as well as adjoining secondary areas, early primates likely retained two or more primary areas and some secondary areas from earlier nonprimate ancestors.

The somatosensory system of early primates was likely similar to those of other mammals at early stages of processing. The system for object recognition included the dorsal-column-medial lemniscus pathways to the ventroposterior nucleus of the somatosensory thalamus. This pathway was more devoted to the tactile receptors of the glabrous skin of the forepaw (hand) than in most mammals, and this skin had larger representations at thalamic and cortical levels. Nevertheless, the thalamic and cortical representations of the face, teeth, and tongue were quite large, with representations of both ipsilateral and contralateral teeth and tongue. The area generally recognized

as area 3b (S1 proper) in anthropoid primates constitutes the homolog of nonprimate S1, although it is common in descriptions of somatosensory cortex in humans to include areas 3a, 3b, 1, and 2 in "S1." In strepsirrhines and some New World monkeys, only one representation of the contralateral body has been found caudal to area 3b, so it is not clear if this region is homologous to area 1 or to both areas 1 and 2. Here we propose that early primates had a single area, corresponding to areas 1 plus 2 of anthropoid primates. Area 3a is a proprioceptive area with inputs from the ventroposterior superior (VPS) nucleus of the somatosensory thalamus. Area 3a, or dysgranular somatosensory cortex was likely retained from early mammals, as were areas S2, PV, and gustatory cortex (G). The ventral somatosensory area (VS), with two divisions, appears to exist in most primates, but has not been identified in other mammals.

Primary motor cortex (M1) and likely one or more premotor areas are found in all primates and were retained from nonprimate ancestors, as M1 and a premotor area are found in rodents and tree shrews (Remple, Reed, Stepniewska, Lyon, & Kaas, 2007). All primates also have dorsal and ventral premotor areas (PMD and PMV), a supplementary motor area (SMA), and two or more cingulate motor areas. The arrangement of these motor fields is remarkably consistent across primate taxa, although premotor areas clearly have functionally distinct subdivisions in some anthropoid primates. Frontal cortex has retained divisions of orbital frontal cortex, while the granular prefrontal cortex is unique to primates (Preuss, 1995). Granular prefrontal cortex is small in strepsirrhines but includes several areas common to primates, including the frontal eye field (FEF). The prefrontal cortex underwent substantial enlargement and elaboration in anthropoid evolution, and areas like the principalis cortex (area 46) of macaques, which mediate higher cortical functions, are most likely anthropoid or catarrhine specializations.

Another region that appears to have changed markedly with the origin of primates is posterior parietal cortex (PPC). In rodents and tree shrews there is little cortex that could be defined as posterior parietal cortex, and visual and somatosensory information reaches motor cortex more directly. However, all primates have a large region of posterior parietal cortex, and part of this cortex has an orderly arrangement of functionally distinct zones that are involved in reaching, defensive, grasping (zones G, D, and R in Fig. 42.7) and other movement sequences (Kaas, Gharbawie, & Stepniewska, 2011). In strepsirrhine galagos, the caudal half of PPC receives most of the visual input, while projecting to the movement-related zones in the rostral half of PPC. Rostral PPC has dense somatosensory inputs, and projects to functionally matched zones in motor and premotor cortex. The organization of PPC is more complex in anthropoid primates, especially in humans where PPC is significantly involved during tool use. PPC also has strong connections with granular prefrontal cortex, and with the higher-order multisensory cortex of the temporal lobe. In primates, these three cortical regions are connected to a common thalamic nucleus, the dorsal pulvinar (also known as the medial pulvinar), and like these cortical regions, the dorsal pulvinar has no obvious homolog in nonprimate mammals (Preuss, 2007).

Brains of Early Anthropoids

The anatomy of early anthropoids suggests that an emphasis on high-acuity, diurnal vision and a reduced emphasis on smell were important in their evolution. The endocasts of the skulls of early anthropoids indicate that the visual cortex of the temporal lobe was more expansive than in strepsirrhines, and overall the brains were considerably larger relative to body size. After the divergence of New World and Old World anthropoids, brain size increased independently in both groups. New World monkeys cover a particularly large range of body and brain size. Since the degree of folding depends on the absolute size of the brain, the smaller-brained New World monkeys like marmosets, owl monkeys, and squirrel monkeys have relatively smooth brains, while the larger species, such as capuchins, spider monkeys, and howler monkeys, have well-fissured brains that resemble (at least superficially) the convoluted brains of Old World monkeys

Both Old World and New World monkeys have all of the areas of the neocortex described for strepsirrhines, as well as some additional areas. Most notably, the somatosensory cortex has been altered so that areas 3a, 3b, 1, and 2 are well differentiated from each other, with each area corresponding to a separate representation of receptors of the contralateral body surface (Kaas, 2007b). Within S1 (area 3b), the proportional representation of body parts varies somewhat across species, so that in some monkeys as much as half of the area is devoted to the face and oral cavity. The hand is also prominently represented, especially in monkeys such as macaques. Some New World monkeys, such as spider monkeys, have a large representation of the hairless (glabrous) ventral skin of their highly tactile, prehensile tail. Marmosets appear to lack a separate area 2, a field that is responsive to tactile stimuli and muscle movements in other monkeys. This may be a consequence of brain size reduction in marmosets, which have evolved smaller brains and bodies from larger ancestors, and transformed their nails back into claws. In all anthropoids, S2 and PV appear to have lost activating inputs from the ventroposterior nucleus of the somatosensory thalamus, and they depend on inputs from areas 3a, 3b, 1, and 2 (Fig. 42.8). S2 and PV receive modulatory inputs from the ventroposterior inferior nucleus. Thus, processing in the somatosensory system became more serial than parallel with the advent of monkeys. Other

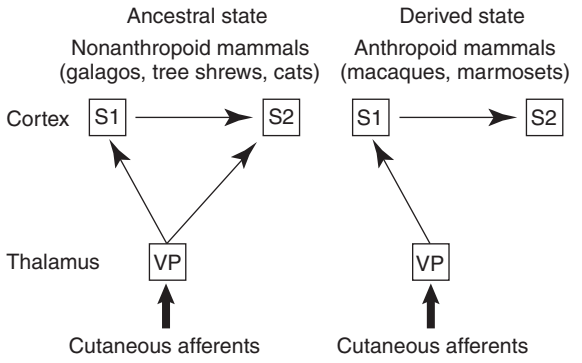


FIGURE 42.8 Somatosensory processing in prosimian primates and anthropoid primates. The processing in anthropoids is serial, rather than parallel and serial. Because the prosimian type is also found in a number of nonprimate mammals, we infer that this is the ancestral state and the anthropoid type is the derived or newly evolved state. The ventroposterior nucleus (VP) of the somatosensory thalamus and the first (S1) and second (S2) somatosensory areas of the cortex are shown.

differences likely exist between anthropoids and strepsirrhines in the somatosensory portions of the lateral sulcus, and more research is needed.

Currently, there is much interest in how the visual cortex of monkeys is subdivided into areas and how these visual areas function in behavior (Orban, Van Essen, & Vanduffel, 2004). Elaborate proposals have been presented, but considerable uncertainty remains. In Old World monkeys, over 30 visual areas have been proposed, and it seems likely that anthropoids in general have more visual areas than strepsirrhines, although the full extent of this difference is not yet clear. The auditory cortex of strepsirrhines has been poorly studied, and it is unclear whether they share the same core-belt-parabelt organization that has been identified in anthropoids. The proposed subdivisions of the motor cortex in strepsirrhines and anthropoids are quite similar, although some of the premotor areas of strepsirrhines have been subdivided into two or three areas in Old World monkeys. Finally, as noted above, the higher-order temporal, parietal, and prefrontal cortex was greatly expanded in anthropoid evolution, and comparative studies of (strepsirrhine) galagos and (anthropoid) macaques indicate that macaques possess a number of areas in prefrontal cortex that galagos lack. The same is likely true of the parietal and temporal cortex, but these regions have not been compared as comprehensively as prefrontal cortex.

Evolution of Hominin Brains

Understanding human brain evolution requires determining what changes took place in the human brain in the 6 to 8 million years since the human and chimpanzee-bonobo lineages diverged. For a long time, nearly all one could say about these changes is that human brains got larger (Box 42.2), reaching about three times the volume of those of chimpanzees and other great apes, although humans are only slightly larger than chimpanzees in adult body size (Fig. 42.9). From the fossil record, we also know that most of this increase took place in the last ~2 my, with the emergence of genus *Homo*, and that there were several species of this genus with brains as large as our own, although only *Homo* sapiens remains today. But the fossil record cannot tell us much about what went on inside the brain, which is what we really want to know if we are to understand the evolution of human-specific behavioral and cognitive abilities, such as language and the ability to represent and reason about abstract or invisible causes, such as mental states (Povinelli & Preuss, 1995; Premack, 2007).

To understand changes in internal organization, we need to compare human and ape brains directly and identify the ways in which they differ. This has posed a great challenge, because neither humans nor apes can be studied with the invasive and terminal techniques so widely used with model animals in the neurosciences. With the development of noninvasive neuroimaging techniques, however, along with the proliferation of antibodies and other molecular tools for probing brain structure in postmortem material, and the rise of comparative genomics, the comparative approach to studying human brain evolution has entered a new and productive era.

Traditionally, it was postulated that the higher-order regions of the frontal, parietal, and temporal lobes (the classical "association" cortex) underwent great expansion in human evolution. This view was recently called into question by comparative structural neuroimaging studies that found that the relative proportions of the frontal, parietal, and temporal lobes are similar in great apes and humans, notwithstanding the much larger absolute size of the human brain. The fact remains, however, that the primary sensory cortical and thalamic structures are of roughly the same absolute size in humans and great apes, while other regions of cortex in humans are much larger in absolute terms. This is consistent with classical accounts of human brain evolution that emphasize the expansion of higher-order regions (Preuss, 2011).

One of the signature specializations of *H. sapiens* is of course language, and it has often been assumed that the evolution of language involved the evolution of specialized language areas in the brain, such as Wernicke's area in the temporoparietal cortex and Broca's area in the frontal lobe. Current evidence, however, strongly suggests that homologues of Wernicke's and Broca's areas are present in apes and monkeys, which in turn suggests that evolution "recruited" these areas and modified them to support language functions. A recent comparative neuroimaging study with diffusion tensor imaging (DTI) has demonstrated that the major pathway between the frontal and temporal language territories (the arcuate fasciculus)

has different connectivity patterns in humans and chimpanzees, with humans showing much stronger connections with the middle temporal gyrus, which in humans is involved in representation of word meanings (Rilling et al., 2008). This is one of several pieces of evidence indicating that the middle temporal cortex expanded greatly in human evolution, displacing the temporal lobe visual areas posteriorly and inferiorly in humans, compared to their location in monkeys.

Since most humans are left-hemisphere dominant for language, it is natural to assume that human brains are more strongly lateralized than those of apes. Surprisingly, apes share some of the gross asymmetries of brain morphology found in humans, including a larger *planum temporale* (the region of superior temporal cortex often identified with Wernicke's area) on the left than the right. Microscopic studies of postmortem tissue, however, indicate that Broca's area, which is larger on the left in humans, is not asymmetrical in size in chimpanzees.

Humans are unusual among mammals in their facility in making and using tools. Functional imaging studies in humans reveal that these activities tap specific, higher-order regions of parietal and frontal cortex (Stout, Passingham, Frith, Apel, & Chaminade, 2011). Direct comparison of brain activity in humans and macaques during tool use indicates that humans have a small territory of left inferior parietal cortex that is selectively responsive when humans view tools that are used to grasp objects, whereas macaque neural activity does not differ between viewing grasping actions of hands and tools (Peeters et al., 2009). It is unclear whether the tool-specific area of humans is a novel area or a modified form of an area present in other primates. A similar issue is raised by the observation (Vanduffel et al., 2002) that the cortex of the intraparietal sulcus is much more responsive to certain types of object movement in humans than in macaques: are these new areas or modified areas? We cannot answer these questions without more detailed comparative studies, with studies of chimpanzees being especially important. In any case, however, it is likely that human cortex is tuned to dimensions or aspects of movement in ways that distinguish humans from at least most other anthropoid primates.

Microscopic studies have also identified specializations of cellular and molecular biology of humans. For example, humans and great apes possess a distinctive type of large, bipolar cell (spindle cells or Von Economo neurons) in anterior cingulate and insular cortex. While present in apes as well as humans, they are especially large and numerous in humans. Similar cells are not found in other anthropoid primates. Interestingly, a similar cell type has evolved independently in whales and elephants, suggesting that this morphology is part of an adaptation for rapid neural communication in large-brained animals (Allman et al., 2011; Watson & Allman, 2007).

The previous example illustrates that evolution can modify areas that are phylogenetically ancient, and not only areas that are evolutionarily new or functionally higher-order. Another example comes from primary visual cortex. Several lines of evidence suggest that apes and humans reduced or lost a projection from the lateral geniculate nucleus to cortical layer 4A, the upper part of cortical layer 4. (Comparative studies indicate that the upper parts of layer 4 of anthropoids, 4A and 4B, are actually parts of layer 3; see Casagrande and Kaas (1994). In addition, humans further modified layer 4A, adding a

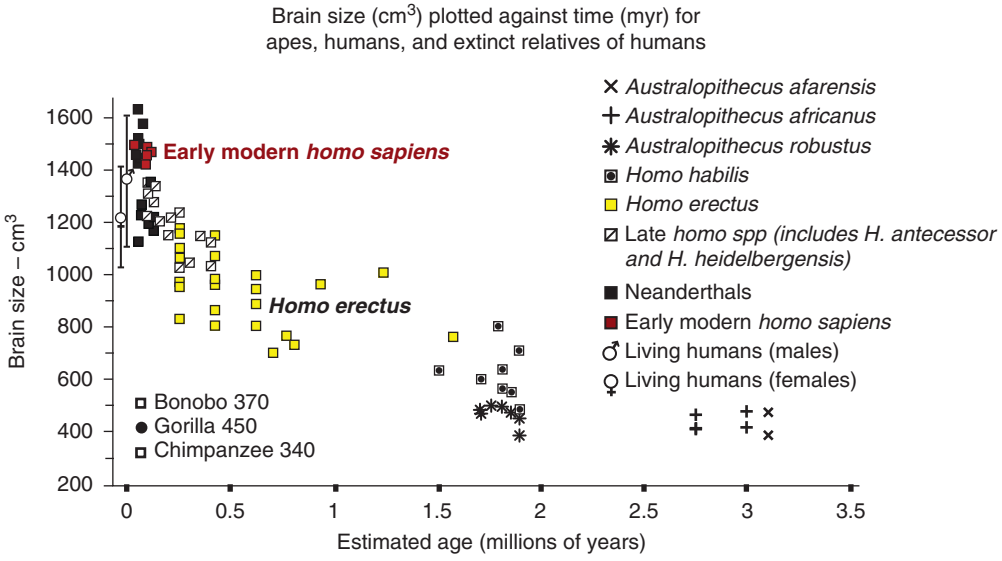


FIGURE 42.9 Evidence for the rapid growth of brains of hominins over the last two million years. The brain sizes of modern chimpanzees and gorillas have been added for comparison. *Modified from McHenry (1994).*

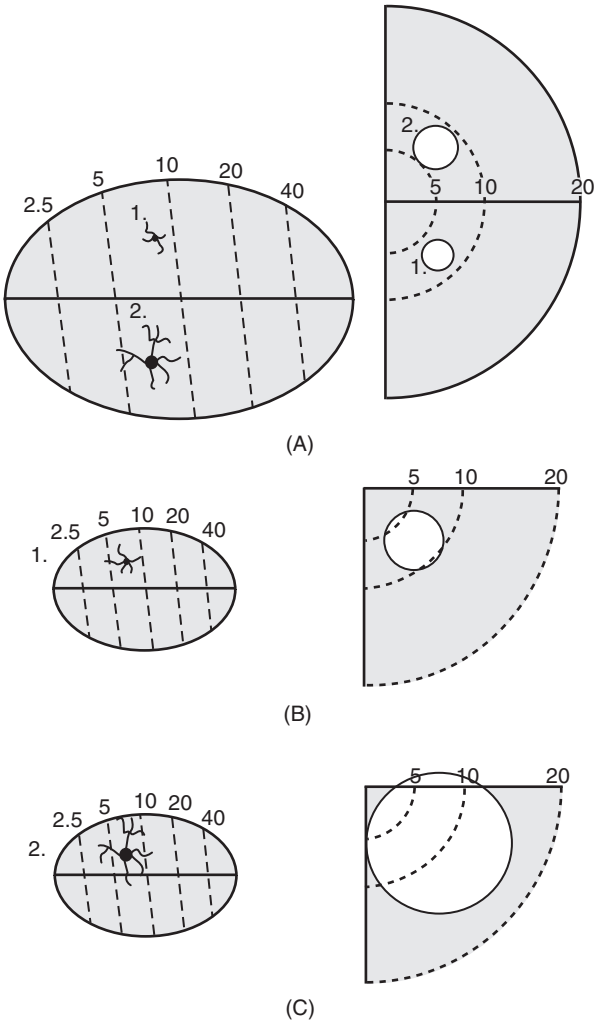


FIGURE 42.10 The effects of varying the horizontal spread of dendritic arbors of neurons in large (A) and small (B and C) visual areas. An increase in arbor size (1 to 2) in a large area (A) produces little change in receptive field size (circles 1 to 2 in the central 20° of the visual hemifield on the right), whereas such a change (B to C) in a small area changes the scope of the receptive field greatly. Thus, the functions of small areas are changed more dramatically by small morphological adjustments. Surface view schematics of retinotopically organized visual areas are on the left, whereas schematics of receptive fields in the visual hemifield and the lower visual quadrant are on the right. *From Kaas (2000).*

novel set of compartments that stain for molecules usually associated with the magnocellular processing stream, the part of the visual system involved in motion analysis (Preuss & Coleman, 2002; Preuss, Qi, & Kaas, 1999). These modifications could influence the functional properties of higher-order areas, including the intraparietal cortex, since V1 is the major source of visual information for the rest of the cortex.

Another new area of active research that provides information about human brain evolution comes from comparative genomics, the goal being to identify evolutionary changes in gene sequences and gene expression relevant to human brain evolution. Although it is commonly assumed that humans and chimpanzees are nearly identical genetically, the most recent genomic comparisons have revealed a plethora of differences, including hundreds of adaptive changes in protein coding sequences; myriad deletions, duplications, and insertions of blocks of DNA; changes in transcription factor binding sites; and differences in the amounts of specific genes and proteins expressed in the brain (Varki & Nelson, 2007). At present, we do not understand what changes in DNA account for the differences in brain size, anatomical organization, and function between humans and other apes, but this is an active area of research. There is, however, intriguing evidence suggesting that the expression of genes involved in synaptic plasticity was modified in human evolution, and that human brains are adapted to sustain unusually high rates of energy metabolism (Cáceres et al., 2003; Fu et al., 2011; Preuss, 2011).

CONCLUSIONS

Based on comparative studies and the fossil record, we conclude that early mammals had small brains with little neocortex and few functional subdivisions (areas or fields) of cortex. Vision was emphasized in the early primates, and the visual cortex in the temporal and occipital lobes was enlarged. These primates also had several unique features of the visual system, including new visual areas such as MT, distinctive kinds of modules in V1 (blobs) and V2 (bands), separate magnocellular and parvocellular layers in the lateral geniculate nucleus, and a representation in the superior colliculus restricted to the contralateral visual hemifield. Several premotor areas were present, whereas the somatosensory system was relatively primitive. Later anthropoid primates had larger brains, more neocortex, and more areas of neocortex. The somatosensory cortex had expanded and included the four strip-like areas on the anterior parietal lobe, areas 3a, 3b, 1, and 2. We know little about possible specializations of the brains of apes. However, over the last 6 million years of evolution in the human lineage, brains increased three to four times in size, due mainly to the enlargement of the cortex. Although the relative proportions of the different cortical lobes remain similar in humans and apes, the available evidence suggests that the cortex did not expand uniformly in human evolution, as the higher-order regions of frontal, temporal, and parietal lobes are much larger in absolute terms in humans than in apes, while human primary sensory areas are about the same absolute size as those of apes. The expansion of nonprimary cortex was probably accompanied by a further increase in the number of cortical areas, modifications leading to functional and anatomical asymmetries in the two cerebral hemispheres, specializations for language and cognition, and larger expanses of prefrontal, parietal,

and temporal cortex. The cortex of the larger brains had many more neurons with a greater proportion of tissue devoted to subcortical white matter connections relative to cell bodies (Herculano-Houzel, Mota, Wong, & Kaas, 2010), and presumably had a higher ratio of local connections to long-distance connections.

Further progress in understanding the evolution of the human brain can be achieved with current methods of investigation. We have the opportunity to learn much about the similarities and differences among the brains of various primates. Neuroscientists have generally concentrated on studies of brain features that are widely shared, but we need to know more about the differences in brain structure and function that make us distinctively human.

References

Aiello, L. C., & Wheeler, P. (1995). The brain and the digestive system in human and primate evolution. *Current Anthropology, 36,* 199–221.

Allman, J. M. (1999). *Evolving brains*. New York, NY: H. W. Freeman.

Allman, J. M., Tetreault, N. A., Hakeem, A. Y., Manaye, K. F., Semendeferi, K., Erwin, J. M., & Hof, P. R. (2011). The von Economo neurons in the frontoinsular and anterior cingulate cortex. *Annals of the New York Academy of Sciences, 1225,* 59–71.

Azevedo, F. A., Carvalho, L. R., Grinberg, L. T., Farfel, J. M., Ferretti, R. E., Leite, R. E., & Herculano-Houzel, S. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology, 513,* 532–541.

Bekkers, J. M., & Stevens, C. F. (1970). Two different ways evolution makes neuron larger. *Progress in Brain Research, 83,* 37–45.

Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Grosshirnrhinde*. Leipzig: Barth. [Reprinted as Brodmann's *Localisation in the cerebral cortex*, (L. J. Garey, Trans. and Ed.). London: Smith-Gordon, 1994.].

Cáceres, M., Lachuer, J., Zapala, M. A., Redmond, J., Kudo, L., Geschwind, D., & Barlow, C. (2003). Elevated gene expression levels distinguish human from non-human primate brains. *Proceedings of the National Academy of Sciences of the United States of America, 100,* 1330–1335.

Casagrande, V. A., & Kaas, J. H. (1994). The afferent, intrinsic, and efferent connections of primary visual cortex in primates. In A. Peters & K. Rockland (Eds.), *Cerebral cortex* (Vol. 10, pp. 201–259). New York, NY: Plenum Press.

Catania, K. C., Lyon, D. C., Mock, O. B., & Kaas, J. H. (1999). Cortical organization in shrews: Evidence from five species. *The Journal of Comparative Neurology, 410,* 55–72.

Changizi, M. A., & Shimojo, S. (2004). Parcellation and area-area connectivity as a function of neocortical size. *Brain, Behavior and Evolution, 66,* 58–98.

Collins, C. E., Airey, D. C., Young, N. A., Leitch, D. B., & Kaas, J. H. (2010). Neuron densities vary across and within cortical areas in primates. *Proceedings of the National Academy of Sciences, 107,* 15927–15932.

de Sousa, A. A., & Wood, B. (2007). The hominin fossil record and the emergence of the modern human central nervous system. In J. H. Kaas & T. M. Preuss (Eds.), *Evolution of nervous systems* (Vol. 3, pp. 292–336). London: Elsevier.

de Sousa, A. A., Sherwood, C. C., Mohlberg, H., Armunts, K., Schleicher, A., MacLeod, C. E., & Zilles, K. (2010). Hominoid visual brain structure volumes and the position of the lunate sulcus. *Journal of Human Evoluation, 58,* 281–292.

Falk, D. (1990). Brain evolution in Homo: The "radiator" theory. *Behavioral and Brain Sciences, 13,* 339–381.

Finlay, B. L., & Brodsky, P. (2007). Cortical evolution as the expression of a program for disproportionate growth and the proliferation of areas. In J. H. Kaas & L. A. Krubitzer (Eds.), *Evolution of nervous systems* (Vol. 3, pp. 73–96). London: Elsevier.

Finlay, B. L., & Darlington, R. B. (1995). Linked regularities in the development and evolution of mammalian brains. *Science, 268,* 1578–1584.

Fu, X., Giavalisco, P., Liu, X., Catchpole, G., Fu, N., Ning, Z. B., & Khaitovich, P. (2011). Rapid metabolic evolution in human prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America, 108,* 6181–6186.

Herculano-Houzel, S. (2009). The human brain in numbers: A linearly scaled-up primate brain. *Frontiers in Human Neuroscience, 3,* 31. Epub 2009 November 9.

Herculano-Houzel, S., Collins, C., Wong, P., & Kaas, J. H. (2007). Cellular scaling rules for primate brains. *Proceedings of the National Academy of Sciences of the United States of America, 104,* 3562–3567.

Herculano-Houzel, S., Mota, B., Wong, P., & Kaas, J. H. (2010). Connectivity driven white matter scaling and folding in primate cerebral cortex. *Proceedings of the National Academy of Sciences of the United States of America, 107,* 19008–19013.

Jerison, H. J. (2007). What fossils tell us about the evolution of the neocortex. In J. H. Kaas & L. A. Krubitzer (Eds.), *Evolution of nervous systems* (Vol. 3, pp. 1–12). London: Elsevier.

Kaas, J. H. (2000). Why is brain size so important: Design problems and solutions as neocortex gets bigger or smaller. *Brain and Mind, 1,* 7–23.

Kaas, J. H. (2007a). Reconstructing the organization of neocortex of the first mammals and subsequent modifications. In J. H. Kaas & L. A. Krubitzer (Eds.), *Evolution of nervous systems* (Vol. 3, pp. 27–48). London: Elsevier.

Kaas, J. H. (2007b). The evolution of sensory and motor systems in primates. In J. H. Kaas & T. M. Preuss (Eds.), *Evolution of nervous systems* (Vol. 4, pp. 35–57). London: Elsevier.

Kaas, J. H. (2011a). Reconstructing the areal organization of the neocortex of the first mammals. *Brain, Behavior and Evolution, 78,* 7–21.

Kaas, J. H. (2011b). The evolution of auditory cortex: The core areas. In J. A. Winer & C. E. Schreiner (Eds.), *The auditory cortex* (pp. 407–427). New York: Springer-Verlag.

Kaas, J. H., & Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences of the United States of America, 97,* 11793–11799.

Kaas, J. H., Gharbawie, O. A., & Stepniewska, I. (2011). The organization and evolution of dorsal stream multisensory motor pathways in primates. *Frontiers in Neuroanatomy, 5,* 34. Epub 2011 June 13.

Martin, RD. (2007). Primate origins: Implications of a cretaceous ancestry. *Folia Primatologica (Basel), 78*(5–6), 277–296.

McHenry, H. M. (1994). Tempo and mode in human evolution. *Proceedings of the National Academy of Sciences of the United States of America, 91,* 6780–6786.

Morwood, M. J., Soejono, R. P., Roberts, R. G., Sutikna, T., Turney, C. S., Westaway, K. E., & Fifield, L. K. (2004). Archaeology and age of a new hominin from Flores in eastern Indonesia. *Nature, 431,* 1087–1091.

Murphy, W. J., Pevzner, P. A., & O'Brien, J. O. (2004). Mammalian phylogenomics comes of age. *Trends in Genetics, 20,* 631–639.

Orban, G. A., Van Essen, D., & Vanduffel, W. (2004). Comparative mapping of higher visual areas in monkeys and humans. *Trends in Cognitive Sciences, 8,* 315–324.

Peeters, R., Simone, L., Nelissen, K., Fabbri-Destro, M., Vanduffel, W., Rizzolatti, G., & Orban, G. A. (2009). The representation of tool use in humans and monkeys: common and uniquely human features. *Journal of Neurosciences, 29,* 11523–11539.

Povinelli, D. J., & Preuss, T. M. (1995). Theory of mind: evolutionary history of a cognitive specialization. *Trends in Neurosciences, 18,* 418–424.

Premack, D. (2007). Human and animal cognition: continuity and discontinuity. *Proceedings of the National Academy of Sciences of the United States of America, 104,* 13861–13867.

Preuss, T. M. (1995). Do rats have prefrontal cortex? The Rose-Woolsey-Akert program reconsidered. *The Journal of Cognitive Neuroscience, 7,* 1–24.

Preuss, T. M. (2007). Evolutionary specializations of primate brain systems. In M. J. Ravosa & M. Dagasto (Eds.), *Primate origins: adaptations and evolution* (pp. 625–675). New York, NY: Springer.

Preuss, T. M. (2011). The human brain: rewired and running hot. *Annals of the New York Academy of Sciences, 1225*(Suppl 1), E182–E191.

Preuss, T. M., & Coleman, G. Q. (2002). Human specific organization of primary visual cortex: Alternating compartments of dense Cat-301 and calbindin immunoreactivity in layer 4A. *Cerebral Cortex, 12,* 671–691.

Preuss, T. M., & Goldman-Rakic, P. S. (1991). Architectonics of the parietal and temporal association cortex in the strepsirhine primate Galago compared to the anthropoid primate *Macaca*. *Journal of Comparative Neurology, 310,* 475–506.

Preuss, T. M., Qi, H., & Kaas, J. H. (1999). Distinctive compartmental organization of human primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 96,* 11601–11606.

Puelles, L. (2011). Pallio-pallial tangential migrations and growth signaling: new scenario for cortical evolution? *Brain, Behavior and Evolution, 78,* 108–127.

Radinsky, L. (1976). Cerebral Clues. *Natural History, 85,* 54–59.

Remple, M. S., Reed, J. L., Stepniewska, I., Lyon, D. C., & Kaas, J. H. (2007). The organization of frontoparietal cortex in the tree shrew (Tupaia belangeri): II Connectional evidence for a frontal-posterior parietal network. *Journal of Comparative Neurology, 501,* 121–149.

Rilling, J. K., Glasser, M. F., Preuss, T. M., Ma, X., Zhao, T., Hu, X., & Behrens, T. E. J. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nature Neuroscience, 11,* 426–428.

Ross, C. (1996). Adaptive explanation for the origins of the Anthropoidea (Primates). *American Journal of Primatology, 40,* 205–230.

Shepherd, G. M. (2011). The microcircuit concept applied to cortical evolution: from three-layer to six-layer cortex. *Frontiers in Neuroanatomy, 5,* 1–15.

Stout, D., Passingham, R., Frith, C., Apel, J., & Chaminade, T. (2011). Technology, expertise and social cognition in human evolution. *European Jouranl of Neuroscience, 33,* 1328–1338.

Vanduffel, W., Fize, D., Peuskens, H., Denys, K., Sunaert, S., Todd, J. T., & Orban, G. A. (2002). Extracting 3D from motion: Differences in human and monkey intraparietal cortex. *Science, 298,* 413–415.

Van Essen, D. C. (2007). Cerebral cortical folding patterns in primates: Why they vary and what they signify. In J. H. Kaas & T. M. Preuss (Eds.), *Evolution of nervous systems* (Vol. 4, pp. 267–289). London: Elsevier.

Varki, A., & Nelson, D. L. (2007). Genomic comparisons of humans and chimpanzees. *Annual Review of Anthropology, 36,* 191–209.

Watson, K. K., & Allman, J. M. (2007). Role of spindle cells in the social cognition of apes and humans. In J. H. Kaas & T. M. Preuss (Eds.), *Evolution of nervous systems* (Vol. 4, pp. 479–484). London: Elsevier.

Wong, P., & Kaas, J. H. (2009). An architectonic study of the neocortex of the short-tailed opossum (Monodelphis domestica). *Brain, Behavior and Evolution, 73,* 206–228.

Wong, P., Collins, C. E., & Kaas, J. H. (2010). Overview of sensory systems of Tarsius. *International Journal of Primatology, 31,* 1002–1031.

Wood, B., & Lonergan, N. (2008). The hominin fossil record: taxa, grades and clades. *Journal of Anatomy, 212,* 354–376.

Suggested Readings

Butler, A. B., & Hodos, W. (1996). *Comparative vertebrate neuroanatomy*. New York, NY: Wiley-Liss.

Deacon, T. W. (1997). *The symbolic species: The co-evolution of language and the brain*. New York, NY: Norton.

Falk, D., & Gibson, K. R. (2001). *Evolutionary anatomy of the primate cerebral cortex*. Cambridge: Cambridge University Press.

Fleagle, J. G. (1999). *Primate adaptation and evolution* (2nd Ed.). San Diego: Academic Press.

Gazzaniga, M. S. (2008). *Human: The science behind what makes us unique* (1st Ed.). New York, NY: Ecco.

Kaas, J. H. (2007). *Evolution of nervous systems* (4 Vols., J. H. Kaas series Ed.). London: Elsevier.

Passingham, R. E. (2008). *What is special about the human brain?*. Oxford: Oxford University Press.

Preuss, T. M. (2009). The cognitive neuroscience of human uniqueness. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences*. (4th Ed., pp. 49–64). Cambridge, MA: MIT Press.

Preuss, T. M. (2010). Reinventing primate neuroscience for the twenty-first century. In M. L. Platt & A. A. Ghazanfar (Eds.), *Primate neuroethology* (pp. 422–453). Oxford: Oxford University Press.

Sherwood, C. C., Subiaul, F., & Zawidzki, T. W. (2008). A natural history of the human mind: Tracing evolutionary changes in brain and cognition. *Journal of Anatomy, 212,* 426–454.

Striedter, G. F. (2005). *Principles of brain evolution*. Sunderland, MA: Sinauer Associates.

Jon H. Kaas and Todd M. Preuss

CHAPTER

43

Cognitive Development and Aging

The human brain has evolved over a very extensive period (see Chapter 42), but an individual human brain develops rapidly over the first few years of life, allowing the nearly helpless human infant to gain control of locomotion, language, and thought. The story of brain development in the early years of infancy through adolescence occupies the first part of this chapter. This is followed by consideration of the cognitive changes that occur during this period of the life span and how these capacities are altered over the course of normal aging. The final parts of the chapter discuss pathological processes that affect cognitive development in childhood and adolescence and can cause devastating cognitive impairment in the elderly.

BRAIN DEVELOPMENT

Basic Concepts of Brain Development

Brain Maturation Progresses Well into Adolescence

Brain development is an organized, predetermined, and highly dynamic multistep process that continues beyond birth into the postnatal period, well into adolescence in humans. Most of our knowledge about human brain development has been obtained through analyses of postmortem specimens and, more recently, powerful neuroimaging techniques. These approaches have not only provided extensive information on how the size and shape of different brain regions change from conception through adulthood in relation to normal cognitive development (Fig. 43.1) but they also have opened unparalleled opportunities for identifying abnormal patterns of development in various psychiatric disorders. The brain grows rapidly in the first 2 years of life, by which time it has achieved 80% of its adult weight. Thereafter, brain growth is slower, reaching over 90% of average adult weight by the sixth year. The development of the cerebral cortical surface proceeds gradually from the flat (lissencephalic) brain of the fetus to the adult gyral pattern at birth. Cortical development in most mammals is broadly divided into two phases. The first is a genetically determined sequence of events occurring *in utero* that can be modified by both the local environment within the fetal nervous system and the maternal environment. The second phase is one that occurs in the perinatal period when the connectivity of the cortex becomes sensitive to patterns of neural activity. Glial processes also critically participate in normal development, supporting cell proliferation, differentiation and guidance, and the formation of myelin sheets around axons. Brain structure size, and developmental trajectories are highly variable and sexually dimorphic.

The Cortex Thickens during Development

Development of the cortical mantle in the human brain was documented extensively in the classic studies of Conel (1939–1963). Longitudinal neuroimaging studies in large samples of children have shown changes in volume of cortical gray matter that are nonlinear and regionally specific (Lenroot & Giedd, 2006). Volume of gray matter increases during preadolescence, with a maximum occurring around 10–12 years of age in the parietal lobe, 11–12 years of age in the frontal lobe, and 16–17 years of age in the temporal lobe, followed by a decline in postadolescence. Although neuroimaging techniques lack the resolution to identify the specific structural modifications that underlie changes in cortical volume during development (i.e., cell number, synapse density, myelination), they do indicate that development of the human cerebral cortex is highly dynamic over time and varies across different cortical areas. Developmental trajectories during

© 2013 Elsevier Inc. All rights reserved.

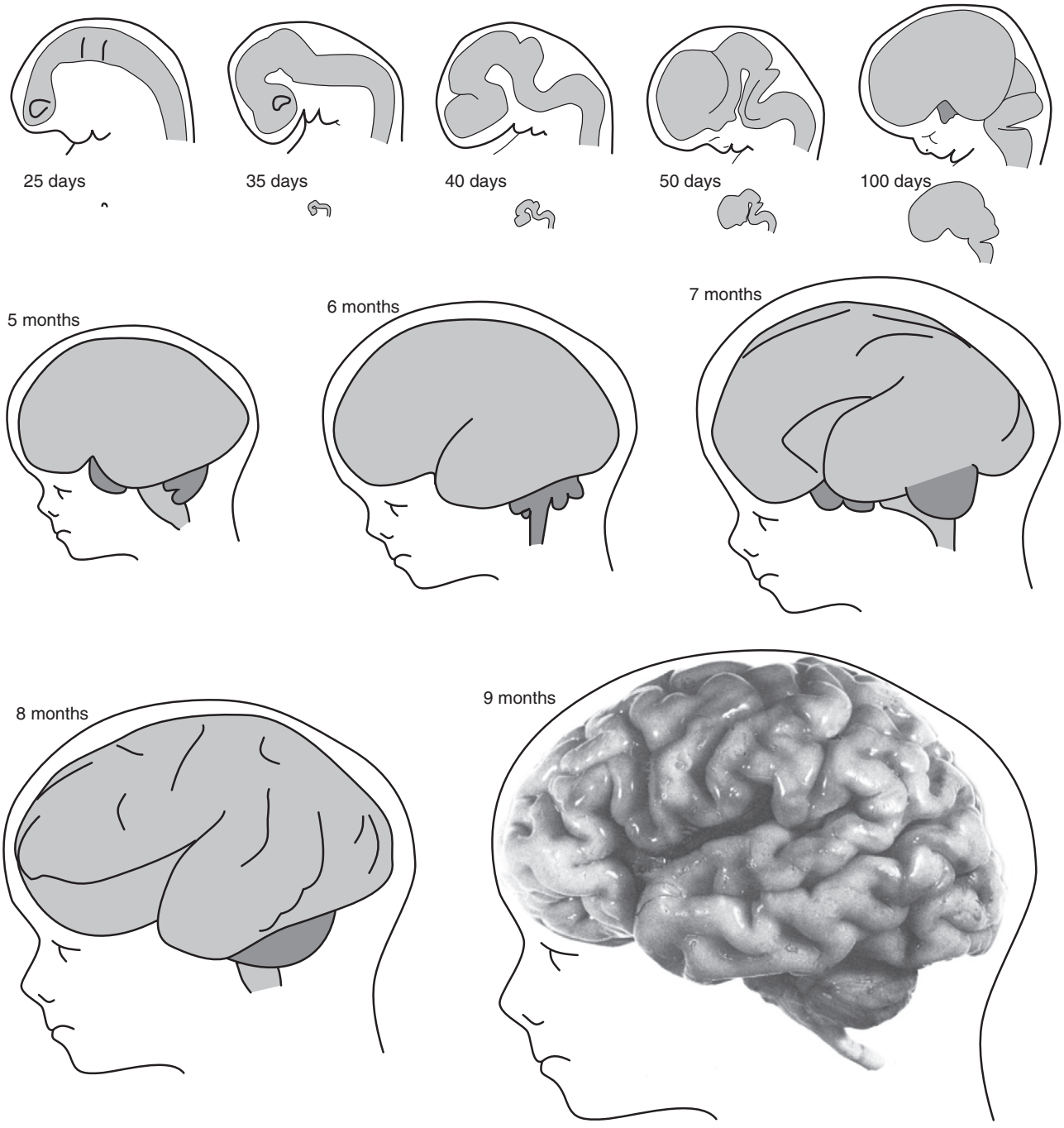


FIGURE 43.1 The size and form of the human brain as it develops through gestation and early infancy.

childhood and adolescence also differ between male and female brains (Lenroot et al., 2007), suggesting that sex steroids play a significant role (Raznahan, Lee et al., 2010).

Dendritic Spine Number Increases During Development

The formation of dendritic spines and the time course of changes in the length and branching patterns of dendrites have been described for visual and frontal cortical areas in humans. Within the visual cortex, peak spine density is achieved around 5 months of age; this number then decreases until adult values are obtained around 21 months of age. Progressive elongation of dendrites occurs up to 24 months. Dendritic development in the visual system reaches mature levels in deep cortical layers earlier than in superficial layers, displaying the "inside-out" pattern of development that is characteristic of neurogenesis and migration (Chapter 15).

Development and maturation of the frontal cortex proceed more slowly. Whereas neuronal density in the primary visual cortex (V1) reaches adult levels by 4–5 months, neuronal density in the frontal cortex fails to reach adult levels even at 7 years of age. Additionally, by 2 years of age, dendritic length in frontal cortex (which is mature by 18–24 months in V1) is only half that found in adults. Left–right asymmetries exist in the dendritic branching patterns of pyramidal neurons within layer V of the inferior frontal and anterior precentral cortex. During the first year, growth is more advanced on the right side, but by 6–8 years of age, the maturation of distal dendrites on the left exceeds that of the right.

Synapses First Increase and Then Decrease in Number

As in the brains of other animals (Chapter 18), the immature human brain contains many more synapses than the mature brain (Huttenlocher, 1990). Within the primary visual cortex, synaptic density increases gradually during late gestation and early postnatal life, and then doubles from 2 to 4 months of age. After age 1, however, there is a decline in synaptic density until adult values (50 to 60% of the maximum) are attained at about 11 years. The time course of the decrease in synaptic density varies across different cortical layers (Fig. 43.2). The decrease does not display the "inside-out" pattern of development; rather, there is a considerable decrease, over time, in the number of synapses in every layer. A postnatal increase and subsequent decline in synaptic density also occurs in the middle frontal gyrus (layer III), but these frontal cortex changes take place over a longer time course than in the primary visual cortex (Fig. 43.3). The maximum density of synapses occurs around 1 year of age (compared to 4 months in the visual cortex), and adult values are not reached until roughly 16 years of age (compared to 7–11 years for the visual cortex).

In summary, these quantitative anatomical results suggest that, in contrast to other animals (Chapter 18), programmed cell death plays only a limited role in human brain development. Synapse elimination in humans and other primates exceeds that observed in many species, however, suggesting that modifications in synaptic connectivity may be particularly important for the development of complex nervous systems. Current findings are consistent with a role for incoming afferent input in selectively stabilizing functional synapses and in eliminating or suppressing inactive contacts. The exuberant connectivity observed during development may also support compensatory processes in cases of early brain trauma, allowing surviving synapses to assume functions normally subserved by damaged inputs.

Postnatal Refinement of Synaptic Contacts Occurs in the Neostriatum and Hippocampus as Well

Although the morphological development of certain subcortical structures, such as the thalamus and cerebellum, is nearly complete at birth, other structures continue to mature into early postnatal life. In the neostriatum, synaptic density increases until the end of the first postnatal month, and changes in neuronal and neuropil morphology are observed until 2–4 months postnatally. Refinement of synaptic contacts proceeds until the end of the first postnatal year when the striatal neurochemical mosaic attains adult characteristics. In the hippocampus, neurons have reached their destination at birth in all CA

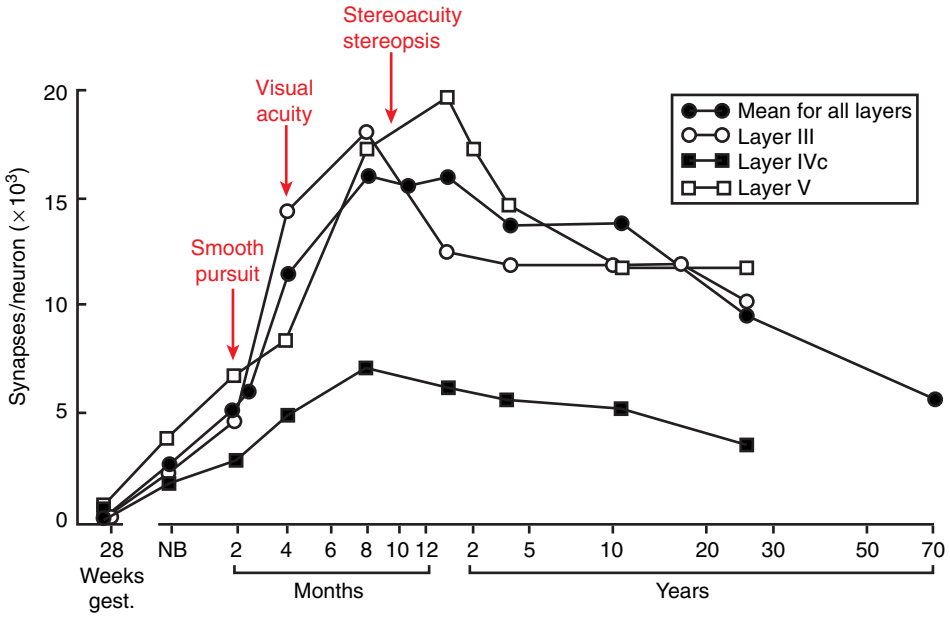


FIGURE 43.2 Variations in the density of synapses of different cortical layers of the human primary visual cortex during development. Arrows point to the emergence of various visual functions in relation to the increase in synapses in the visual cortex. NB, newborn. *Adapted with permission from Huttenlocher and De Courten (1987).*

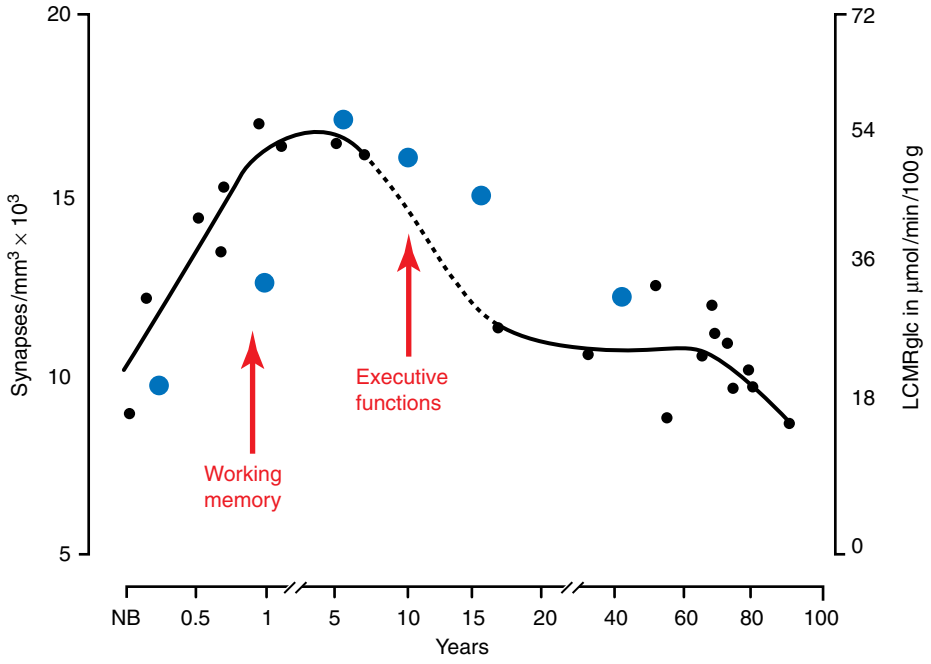


FIGURE 43.3 Variations in the human prefrontal cortex during development: Density of synapses in layer III of the medial frontal gyrus (black circles) and resting glucose uptake (LCMRglc) in the frontal cortex (blue circles). Arrows point to the approximate periods of emergence of various prefrontal cortex functions. NB, newborn. *Adapted with permission from Huttenlocker (1990) and Chugani, Phelps, and Mazziotta (1993).*

fields, but immature cells continue to accrue within the dentate gyrus throughout the first year of life. In addition, continued dendritic remodeling in the hippocampus has been observed as late as the 5th postnatal year in humans (Seress, 2008). Myelination in the subicular and presubicular regions, a key relay zone between the hippocampus and many cortical areas, continues through adolescence and adulthood (Benes, Turtle, Khan, & Farol, 1994) with female subjects showing a greater extent of myelination between the ages of 6 and 29 than males. These changes are accompanied by corresponding changes in the volume of the hippocampus during early development (Giedd et al., 1996).

White Matter Increases Throughout Childhood and Adolescence

In contrast to the inverted U shape of gray matter development, white matter maturation in the brain continues long after the gray matter has reached its definitive volume (Lenroot & Giedd, 2006; Klingberg, 2008). In general, myelination progresses from caudal to cephalic. Thus, the brainstem myelinates prior to the cerebellum and basal ganglia. Similarly, the cerebellum and basal ganglia myelinate prior to the cerebral hemispheres. Sensory and motor systems display mature myelination within the first 2 years of life. However, nonspecific thalamic radiations do not reach mature levels until 5–7 years of age, and intracortical fibers continue their myelination processes well into the third decade of life. Numerous morphological studies demonstrate that the corpus callosum is uniformly thin in the first month postnatally. The genu and then the splenium undergo rapid growth spurts during the next 5 months, resulting in their characteristic "bulbous" appearance.

Cerebral Metabolic Rate Increases and then Decreases after Birth

Increases in neuronal activity have been linked to increases in cerebral metabolism, as measured by neuroimaging techniques. Data show substantial subcortical activation in newborns, notably in the thalamus and phylogenetically old portions of the cerebellum, but little activation of the cerebral cortex. However, over the first 3–4 years of life, cortical metabolic rate increases until it reaches levels twice those observed in adults. After 4 years of age, metabolic activity declines gradually to adult levels at around age 15. The time course of the rise and decline of activity, as revealed by positron emission tomography (PET), varies across cortical regions but parallels the rise and decline in the number of synapses in these areas, suggesting that the exuberant cortical synapses observed during development are metabolically active. As these synapses are eliminated, metabolic activity decreases.

Dorsolateral Prefrontal Cortex Is One of the Last Brain Regions to Develop

Refinement in the circuit organization of the dorsolateral prefrontal cortex, like that of the hippocampal formation (see earlier discussion), continues until quite late in development. In the case of the dorsolateral prefrontal cortex, a gradual decline in synaptic density continues well into late adolescence and suggests that connectivity

in this cortical region undergoes substantial change until 15–20 years of age. These changes are accompanied by a decrease in metabolic rate during the same period (Fig. 43.3). It is believed that changes in the neuromodulatory effects of dopamine may influence the adolescent refinement of excitatory and inhibitory inputs to layer III pyramidal neurons and that dopamine may have a particularly strong influence on cortical information processing around the time of puberty (for the relevance of this late prefrontal cortex development in the emergence of psychopathologic states, such as schizophrenia, see the section on pathological processes).

Summary

Structural, metabolic, and physiological indices of human brain development all point to a long postnatal time course that displays considerable variability from region to region and from system to system. Thus, while postnatal development in the hippocampus proceeds until approximately 4–5 years of age, it continues until 7–11 years for the visual cortex and until around 16 years of age for the frontal cortical areas. The dorsolateral portion of the prefrontal cortex appears to be one of the last regions of the brain to develop. Interestingly, the development of myelin has been observed as late as the third to fourth decade of life for fibers within the association areas of the neocortex. Finally, brain structure, size and developmental trajectories are sexually dimorphic.

COGNITIVE DEVELOPMENT AND AGING: A LIFE SPAN PERSPECTIVE

Investigators are now attempting to link the types of neural changes described in the preceding section to the maturation of sensory, behavioral, and cognitive capacities in infants and children. Such studies are limited in number. Until recently, the predominant view was that postnatal development of the cortex is largely intrinsically determined. However, new studies of both animals and humans have revealed a central role for extrinsic factors in shaping the organization of neural systems and in permitting recovery from brain damage. Indeed, calculations showing that information in the genome is not sufficient to specify the connectivity of the brain, together with evidence for the long-lasting existence of transient, redundant connections in primates, suggest that neural changes under the influence of environmental input play a significant and persistent role in the development of functional specificity in the human brain. Thus, genetic and environmental influences are intimately intertwined in guiding brain development (Lenroot et al., 2009).

Cognitive Development in the First Years of Life

Anatomical Development of the Visual System in Infants Is Linked to Functional Development

Several parallels have been noted between anatomical changes and the emergence of visual system function in humans. The visual abilities of the newborn have been linked to subcortical structures that show mature anatomy and high metabolic rates during this period. The limited visual abilities of the newborn are augmented by the appearance of smooth pursuit tracking around 6–8 weeks. The emergence of this capacity may be related to the maturation of cortical layers IV, V, and VI in V1 (Fig. 43.2). These layers connect magnocellular afferents from the lateral geniculate nucleus and the middle temporal (MT) region, a pathway important for the detection of visual movement. Visual acuity and visual alertness increase dramatically around 4 months of age, when the volume of visual cortex reaches adult levels and the highest density of synapses is present. Shortly thereafter, cortically mediated binocular interactions become apparent. These interactions, which include stereoacuity, binocular summation of the light reflex, and stereopsis (binocular depth perception), appear in the same time frame as maturation of the middle cortical layers and rapid synaptogenesis in V1. This period of rapid growth appears to be a time of increased vulnerability to altered afferent input. Strabismic amblyopia and amblyopia due to the absence of patterned input (i.e., centrally mediated visual impairments) are reported to occur at this age unless corrected early. The time period during which visual impairments can be corrected is different for different visual functions. For example, correction for an absent lens (aphakia) due to cataracts may be completely effective only when performed prior to 2 months of age—that is, just prior to the onset of exuberant synaptogenesis in V1. However, very high synaptic density persists to at least the age of 4 years in V1. The presence of this extensive connectivity may account for the ability to recover from amblyopia (with forced use of the strabismic eye) during this time. If the disruption of binocular convergence by strabismus is not corrected within the first year of life, the ability to see stereoscopically may not develop even though the development of acuity and contrast sensitivity is normal. Thus, there may be separate critical periods (see Chapter 21) for the development of resolution acuity and stereopsis. Because the parvocellular system is thought to underlie acuity, whereas disparity detection is mediated by the magnocellular system, the magnocellular system may be more modifiable by environmental input than the parvocellular system. This differential sensitivity may be due to the slower maturation of the magnocellular system and/or to differences in the number of exuberant synapses within these systems.

Studies of individuals born deaf and blind suggest that in humans, as in other animals, there is a time period when cortical areas that normally process information from the deprived modality can be reorganized to process information from intact modalities. From the alteration of visual functions seen in individuals born deaf, it appears that areas of the auditory cortex have been recruited for visual function. Deaf subjects, by comparison, show abnormally strong electrical responses to peripheral visual stimuli. These responses are not found in normally hearing individuals, even those exposed to sign language by deaf parents. The prolonged persistence of exuberant cortico-cortico connections may provide the substrate for such cortical reorganization.

Development of the Ventral Visual Stream and Object Representation

A multitude of visual cortical areas located in the striate and extrastriate cortex contribute to different aspects of visual processing (see Chapter 44). These visual areas have been divided into two visual processing systems, one coursing ventrally from the striate and extrastriate areas to the ventrolateral temporal cortex and the other coursing dorsally from the striate and extrastriate cortex to the parietal cortex. The dorsal stream mediates spatial processing associated with attention to movement and location, whereas the ventral stream is involved primarily in processing information about patterns and objects. Both streams project rostrally, each reaching common and adjacent areas of the prefrontal cortex.

Major functions of the ventral stream include the ability both to segment a pattern into a set of constituent parts (local level of processing) and to integrate those parts into a coherent whole (global level of processing). In adults, systematic differences exist in the distribution of global and local levels of processing within the brain, such that global levels of processing are associated with the right posterior temporal region and local levels of processing with the left posterior temporal regions. Around 4 months of age, infants show hemispheric differences for global and local processing that are similar to adults. Thus, at this age, children segment out well-formed units and use simple associations to organize these units into a configuration. With further maturation, children show changes in the way they decompose objects into parts and in the relations they use to organize the parts. These changes may reflect the protracted development of the posterior temporal cortical areas, a finding consistent with ERP studies in human infants indicating functional activation within the temporal cortical areas around 6 months of age. The latter observation parallels the emergence, around the same age, of object representation and visual rule learning (Alvarado & Bachevalier, 2000).

Development of the Dorsal Visual Stream and Spatial Attention

A variety of spatial processes have been associated with activation of the dorsal visual pathway, such as the processing of information about spatial attention and spatial location (see Chapter 45). Considerable clinical and experimental evidence shows that the posterior parietal lobe plays a crucial role in the ability to shift attention to different spatial locations. By at least 6 months of age, infants, like adults, show facilitated stimulus detection as a function of prior cueing—that is, they take less time to detect an object if it is presented at a location that has been cued previously. They also demonstrate, as do adults, quicker target detection when the interval between the cue and target is short than when this interval is long. While this basic attentional response may be robust as early as 4 months of age, the timing parameters that elicit the response change with development, suggesting that further maturational processes occur in the parietal cortex.

Precursors of Declarative Memory Emerge Early in Development

Looking preference has been used to study the emergence of recognition memory ability in both human infants and monkeys (Bachevalier & Vargha-Khadem, 2005). In both species, the most common procedure is to first present the infant with an attractive stimulus for a period of time (generally 20 to 30 s). After an intervening delay that can vary from a few seconds to many hours, the familiar stimulus is shown together with a new item. Human infants as young as 1 day old display a strong preference for looking at the novel object after delays as long as 24 h, demonstrating that they recognize the previously presented item as familiar. The early emergence of recognition memory is also observed in infant monkeys and, in this species, surgical removal of the hippocampal formation and adjacent tissue before the age of 3 weeks disrupts this ability only when the hippocampus normally becomes fully mature (around 1 year of age in monkeys). Thus, early developing recognition memory processes are likely mediated by allocortical areas (phylogenetically older cortex) of the medial temporal lobe, such as the perirhinal cortex and parahippocampal gyrus. These cortical areas are known to develop earlier than the hippocampus (Zeamer & Bachevalier, 2010). The capacity for recall over long periods of time, as measured by nonverbal deferred imitation tasks, also emerges by the end of first year in humans and is impaired after early hippocampal injury (Bauer, 2002; de Haan, Mishkin, Baldeweg, & Vargha-Khadem, 2006).

The early emergence and continued improvement of recognition and recall abilities is consistent with what is known about maturation of the hippocampal formation in both humans and monkeys (see earlier discussion). In both species, the hippocampal formation is almost

adult-like at birth, although there are a number of postnatal morphological refinements that continue until the end of the first year in monkeys and until about 4–5 years of age in humans. These protracted modifications in hippocampal circuitry, together with the further maturation of cortical areas in the temporal and prefrontal lobes, may provide a basis for the similarly late elaboration of certain capacities, including spatial and relational memory (see Chapter 48). Evidence that contributions of the medial temporal lobe and prefrontal cortex to declarative memory follow distinct developmental trajectories has also been documented by functional neuroimaging (Ofen et al., 2007).

Orientation to Faces, Recognition of Mother's Face, and Imitation of Facial Expressions Are Present at Birth

The human face is a powerful visual source of social information, and the development of face-processing abilities appears to follow a delayed postnatal development. Newborns preferentially orient to faces, recognize the mother and other familiar faces, and imitate facial expressions. These precocious tendencies appear to be mediated by a subcortical retinotectal pathway. For example, newborns display preferences selectively for moving stimuli presented in the peripheral visual field (i.e., under conditions that engage the subcortical systems), but not when they are displayed in the central visual field.

The fact that newborns just hours to days after birth look longer at the mother's face than at a stranger's (even when cues from her smell and voice are eliminated) indicates that, from very early on, there is a mechanism capable of learning about individual faces based on experience. This early learning system may be mediated by early developing allocortical areas known to be involved in memory (see earlier discussion; Shaw et al., 2008). Nevertheless, the face-processing abilities of the newborn are qualitatively different than the sophisticated capacities of adults, and none of these tendencies are necessarily the direct precursors of the cortically mediated, adult face-processing system (see Chapter 44). Instead, these responses might serve the purpose of providing input to developing cortical circuits that will at some later time functionally emerge to mediate face processing.

At 3 Months of Age Infants Are Able to Form Categories of Visual Stimuli

A marked change in infants' visual attention to faces occurs at approximately 8 weeks of age. At this time, infants' preferential following of peripheral moving faces declines, and a preference emerges for fixating faces compared to other patterns presented in the central visual field. This behavioral change is thought to be due to the functional development of visual cortical pathways that inhibit the preferential following response and mediate the new preferential fixation response. At 2 months of age, infants become more sensitive to the internal facial features of static faces. In addition, within the face, the eyes are a more salient feature than the nose or mouth, but where the eyes are located is immaterial to the babies' preference for face-like drawings. At roughly the same age, infants begin to relate information between individual faces and to form categories; that is, they perceive visual stimuli with comparable features as being more similar than visual stimuli with different features. They also exhibit inversion effects (i.e., recognition of upright faces is easier than recognition of inverted faces). The change between 1 and 3 months has been associated with the functional development of temporal cortical areas and their connections with the hippocampus and adjacent structures. However, at this early age, the change may not be specific to faces, as 3-month-old infants can form perceptual categories of a variety of other complex categories, such as tables and trees.

During childhood, the child's prototype of the face becomes more tuned to the types of faces that the child sees most often, and with this may emerge race effects and other species effects seen in adult face processing. Other race effects refer to the ability to distinguish among faces of one's own race more quickly than faces of other races. Other species effects refer to the ability to distinguish between human faces more quickly than faces of other species. These last changes may be based on experience due to increased exposure to faces, the number and type of features children attend to and encode, and the maturation of a posterior temporal cortical area (the fusiform area) that is known to be involved in face processing in adults.

Fear of Unfamiliar Faces Emerges between 7 and 9 Months

Although infants are able to discriminate among some facial expressions of emotion by 3 months of age, they do have difficulty discriminating between sadness and surprise. It is only by 7 months of age that infants, like adults, show categorical perception of facial expressions. For example, with pictures of faces in which emotional expressions are progressively degraded from happy to fear, human infants and adults exhibit more accurate discrimination for pictures of faces that cross emotional categories (happy versus fear) than for pictures of faces within the same emotional category, despite equal physical differences in the pictures. With this increased ability to categorize faces, and the increased exposure to familiar faces, two remarkable events occur between 7–9 months of age: the emergence of fear for unfamiliar people and anxiety during temporary separation from the caretaker. Distress under both conditions must involve more than

the ability to discriminate strangers from parents because 3 month-olds can make this discrimination, but do not show fear reactions.

One speculation concerning stranger and separation fear is that improvements in working memory, and an enhanced ability to retrieve memories of a past event, are required for the appearance of these reactions. This change appears to be associated with the maturation of the orbital frontal cortex (Schore, 1996), a region of the prefrontal cortex intimately interconnected with limbic areas in the temporal lobe (temporal pole cortex and amygdala), with subcortical drive centers in the hypothalamus, and with dopamine neurons in reward centers in the ventral tegmental area. By the end of the first year, the initial phase of orbital frontal maturation is achieved and allows for developmental advances that enable the individual to react to situations on the basis of stored representations, rather than on information immediately present in the environment. In the case of the orbital frontal cortex, this capacity applies specifically to socioemotional information. Indeed, by 10 months, infants are first able to construct and store abstract prototypes of human visual facial patterns, and can use these prototypes to evaluate novel information.

In an illuminating experiment, a preferential looking task was used to habituate 10-month-old infants to pictures of faces. One condition presented faces of different females, and in the other condition, the same female face was presented in various poses. After infants habituated to the stimuli (i.e., decreased their looking), they were given two test trials: one with a face of a familiar female and another with a face of a novel female. When infants were previously habituated using different female faces, they spent equivalent amounts of time gazing at familiar and novel faces on the test trials. However, when infants were habituated to the same face with different poses, they generalized their looking response to the familiar face but dishabituated (or looked longer) at the novel face. This pattern of results suggests that infants can abstract relevant categorical information by 10 months of age.

Changes in Visual Search Are Related to Development of the Prefrontal Cortex

Infants younger than 7 or 8 months of age will not uncover a hidden object. If a cloth is thrown over a toy while an infant of 5 or 6 months is reaching for it, the infant will withdraw his or her hand and stop reaching. By 7 or 8 months, most infants who watch an object being hidden can retrieve it. However, if the infant then watches as the object is hidden at a second location, most infants of 7–8 months search for the object at its first hiding location, which is now empty. Jean Piaget called this the A-not-B error because the infant is correct at the first hiding place (A) but not at the second (B). It is rare to see this error when there is no delay between when the object is hidden and when the infant is allowed to reach. However, a delay as brief as 1, 2, or 3 s is sufficient to produce the error in infants 7–8 months of age. As they grow older, the error is still seen, but only if the delay between hiding and retrieval is increased. The A-not-B error suggests that during the second half of the first year the brain has matured sufficiently to enable the infant to hold a representation of the object's location in mind (working memory) for a few seconds (see Chapter 50) or to understand the relationship of (a) their previous action of retrieving the object, (b) the subsequent hiding of the object in a different location (even though they observed the hiding), and (c) the object's present whereabouts (Diamond, 1990). The enhanced working memory observed in the 7- to 10-month-old infant parallels the protracted maturation and refinement of dorsolateral prefrontal cortex circuitry (Fig. 43.3). After 12 months of age, it is difficult to elicit the A-not-B error.

Language Undergoes Rapid but Prolonged Development

In humans, the prolonged structural, metabolic, and neurophysiological maturation of "association" cortical areas, which continues well into adolescence, provides the substrate for the panoply of higher cognitive functions that continue to develop during this time. Attempts have been made to link the very rapid development of speech and language skills over the first 3 years of life to these general changes. This goal has remained elusive, probably because so many aspects of the brain and behavior are changing together. Moreover, key elements of language appear to be processed at cellular and synaptic levels of organization that are difficult to investigate in humans with even the most sophisticated techniques available.

There is wide agreement that language is strongly dependent on structures within the left perisylvian region (see Chapter 49). Very early on (by 28 weeks of gestation), structural asymmetries appear between the temporal lobes; these may provide the substrate for the functional asymmetries that appear later. The rapid and early acquisition of phonological information and speech production and comprehension, and the subsequent burst in size of the vocabulary, may be linked to the rapid rise in the number of synapses and the marked increases in cortical metabolism that occur during the second year of life. Also, the persistence of large numbers of exuberant synapses through adolescence may provide the anatomical substrate for prolonged neural plasticity and recovery of language skills following cortical damage in the first decade of life. It is well established that language skills can display considerable recovery following large lesions involving the left hemisphere during the first 7–10 years of life.

The impressive language recovery in children who have undergone left hemisphere removal is even more striking in light of the enduring impairment seen in specifically language-impaired children or the reading difficulties encountered by dyslexics, in whom macroscopic aspects of brain structure are basically normal. These findings underscore the importance of characterizing microscopic structural aspects of the brain and the functional organization of the brain in relation to processing.

The prolonged time course of development that may confer plasticity on the immature brain appears to be characterized by optimal or critical periods (see Chapter 21) for language acquisition. Several studies report that both first and second language acquisitions are impaired, and cerebral organization is altered, when language is acquired after the first decade of life. Moreover, as has been observed for vision (Fig. 43.2), different aspects of language appear to display different critical periods. Vocabulary items can be acquired long past the first decade of life, but the grammatical rules of a language appear to be acquired most readily before the age of 10. Along with other evidence, this pattern suggests that different neural systems, with differing developmental time courses, mediate these various aspects of language.

Summary

Cognitive development parallels some of the most important changes in brain development. Thus, the visual abilities of the newborn are limited and control by subcortical systems predominates. With the maturation of visual cortical areas in the first few postnatal months, visual acuity, visual alertness, and stereopsis develop. With further development of visual cortical areas within the temporal lobe, an infant's ability to recognize faces using internal facial features, mostly the eye region, emerges around 3 months of age. However, it is only by 7–9 months of age that infants develop fear of unfamiliar people. It is believed that this behavior is associated with the development, around this age, of the orbital frontal cortex, a cortical area that enables the infant to react to situations on the basis of stored representations (working memory). The enhanced spatial working memory abilities observed by the middle of the first year parallel the protracted maturation and refinement of the dorsolateral prefrontal circuitry. Finally, the maturation of association areas of the cortex, which continues well into adolescence, provides the substrate for the development of higher cognitive functions, such as language. In the future, it will be possible to analyze the neural basis of cognition in increasing detail by using high-resolution methods for imaging brain structure and function in normally developing children and children with specific structural or functional deficits.

COGNITIVE AGING

Average Human Life Expectancy has Increased Dramatically

The world's population is growing older. Whereas life expectancy in the United States was approximately 50 years in 1900, infant mortality rates decreased dramatically during the last century, and combined with advances in disease prevention and treatment, current life expectancy is approaching 80 years in many industrialized countries. A significant demographic trend is that the percentage of the population over 65 years of age is projected to more than double between the years 2000 and 2030, with particularly rapid growth among those over 85. Neurodegenerative disorders such as Alzheimer's disease are among the most devastating consequences of growing older (see section on Dementia, this chapter). There is increasing recognition, however, that many otherwise healthy aged individuals experience deficits in memory and other aspects of cognitive function that, although less severe, pose a significant risk to independence and the quality of life. The following sections outline key concepts from the study of normal cognitive aging in the absence of frank disease.

Effects of Normal Aging on Memory Are Variable

Cognition encompasses a variety of dissociable capacities, including perception, attention, executive function, and memory (Section VII, this volume), mediated by partially distinct neural systems. On the basis of this organization, neuropsychological research documenting the specific types and severity of cognitive deficits that accompany aging can provide important clues about the brain systems that mediate dysfunction. This strategy has been widely exploited to examine the effects of normal aging on learning and memory, where a rich background is available concerning the characteristics of impairment that follow damage to selective brain regions in young adults.

A key concept to emerge from neuropsychological studies of aging is that the status of memory mediated by the hippocampal formation and related medial temporal lobe structures varies substantially among older individuals (Fig. 43.4). In humans and animal models, memory ability is continuously distributed across a broad range such that some aged individuals exhibit considerable impairment while others—at the same chronological age—perform as well as young adults. Long considered a complicating factor in gerontological research, this variability suggests the important conclusion that marked deterioration in memory is not an inescapable consequence of growing older and that the functional integrity of the medial temporal lobe system is preserved in a substantial proportion of aged individuals. The next section considers findings

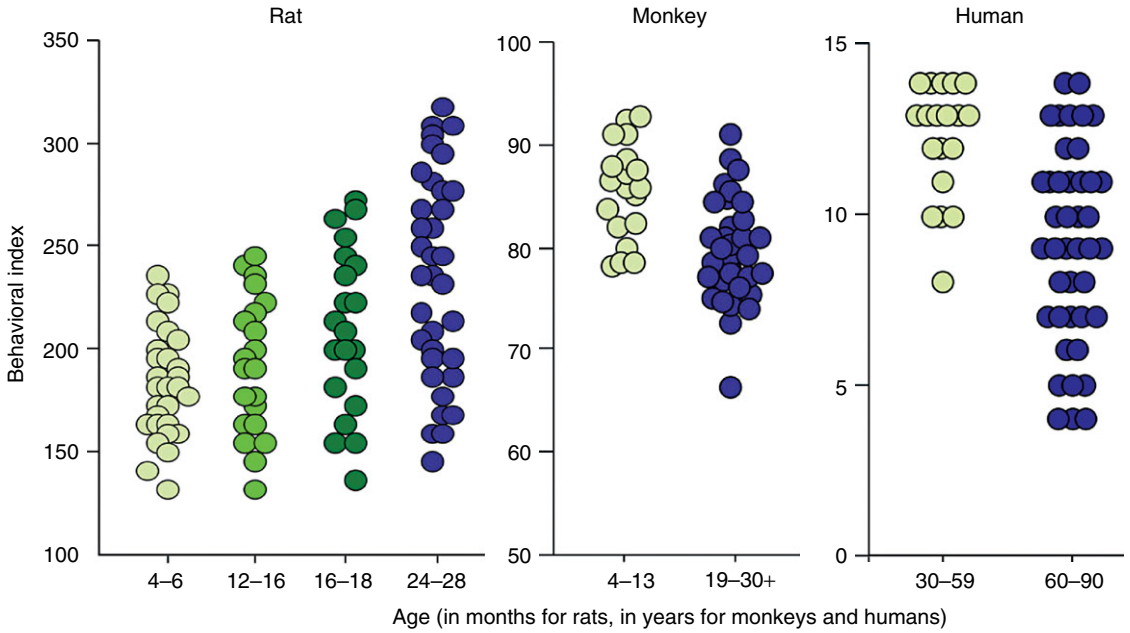


FIGURE 43.4 The status of memory varies considerably across aged individuals. The data comprise a summary measure of spatial memory in rats (courtesy of M. Gallagher, Johns Hopkins University), object recognition memory in monkeys (courtesy of P. Rapp, National Institue on Aging), and delayed recall in humans (courtesy of M. Albert, Johns Hopkins School of Medicine). For rats, low scores represent better memory, and high values reflect better performance in monkeys and humans. Symbols signify scores for individual subjects. Note that in all species the status of memory among the aged is distributed across a broad range, from individuals that score on a par with the best young adults to other individuals that exhibit substantial impairment.

from animal models supporting this perspective, and that have guided efforts to identify the neurobiological alterations reponsible for individual differences in normal cognitive aging.

Animal Models of Normal Neurocognitive Aging

As described in Chapter 48, normal adult rats solve the spatial, or "place," version of the Morris water maze by learning and remembering the escape location in relation to cues surrounding the testing apparatus, and this capacity requires the hippocampal formation. Guided by this background, many laboratories have used the Morris water maze and other spatial tasks to examine the effects of aging on learning and memory mediated by the hippocampal system (reviewed in Rosenzweig and Barnes, 2003).

Aged rats learn the hidden platform location in the place version of the water maze more slowly than young adults. Although these findings are consistent with a possible hippocampal contribution to cognitive aging, an alternative account is that declining visual acuity, motor function, or other impairments give rise to performance deficits in older subjects. This possibility has been tested using control procedures that make many of the same sensory, motor, and motivational demands as the hidden platform task, but that lack an explicit spatial learning component. In a common task variant, for example, rats swim to a platform that protrudes above the surface of the water, providing a salient visual goal to guide navigation. Because the cued platform location is varied randomly across trials, learning and memory for spatial information are not required. Many aged rats acquire this task normally, demonstrating that they can swim proficiently, are motivated to escape, and retain sufficient visual acuity to support accurate cue-directed navigation. Thus, whereas some aspects of motor and visual function are indeed vulnerable to aging, these impairments fail to explain the deficits aged rats exhibit in the spatial version of the water maze. Instead, the overall profile observed in aging—impaired spatial learning and memory against a background of intact goal-approach learning—is qualitatively similar to the effects of damage to the hippocampal system in young subjects. Other tasks that require hippocampal integrity are also sensitive, confirming that learning and memory mediated by this system are susceptible to age-related decline (Rosenzweig & Barnes, 2003).

As noted earlier, research in animal models has revealed substantial individual variability in the cognitive outcome of aging. Figure 43.4, for example, illustrates individual learning scores for large numbers of young, middle-aged, and aged rats tested in the spatial version of the water maze. The performance measure in this case reflects distance from the escape platform, averaged over multiple trials, and accordingly, low scores represent better learning, with searching focused near the escape

location. Similar to the distribution of memory scores for people (Fig. 43.4), spatial learning capacity among aged rats varies across a broad range such that some perform as well as even the most proficient young subjects. By comparison, approximately half of the older animals exhibit marked deficits, scoring outside the range of the young adult group.

Variability in the cognitive effects of aging has been examined in nonhuman primates as well, taking advantage of testing procedures adopted from studies on the effects of medial temporal lobe damage in young adult monkeys (see Chapter 48). In the delayed nonmatching to sample task, subjects are tested for their ability to recognize a visual stimulus after a delay ranging from seconds to many minutes or more. Monkeys with sufficiently extensive damage to any component of the medial temporal lobe memory system display significant deficits on this task, particularly under conditions of increased memory demand, with long retention intervals. Recognition memory declines during aging in the monkey, and as in humans and rats, the degree of impairment varies substantially across individual subjects (Fig. 43.4). As discussed in the next section, this variability makes it possible to ask what neurobiological alterations distinguish the brains of individuals with age-related memory impairment from other, age-matched subjects with preserved function. The aim by this approach is to identify changes in the structure and physiology of the aged brain that are specifically coupled to individual differences in the cognitive outcome of normal aging.

Neurobiology of Age-Related Memory Decline

Age-Related Learning and Memory Impairment Is Associated with Compromised Hippocampal Physiology

Behavioral studies suggest that the functional integrity of the hippocampal system declines during aging. Electrophysiological investigations extend these observations by directly assessing the computational processing functions of the aged hippocampus. Models of learning-related cellular plasticity, such as hippocampal long-term potentiation and depression (LTP and LTD; see Chapter 47), establish a useful background against which to explore the effects of aging. A prominent theme is that aging influences hippocampal physiology in a highly selective manner, sparing many aspects of function (reviewed in Burke and Barnes, 2006: and Wilson, Gallagher, Eichenbaum, and Tanila, 2006). Parameters that exhibit little or no change in older subjects include the resting potential, input resistance, and the amplitude and duration of evoked action potentials of principal hippocampal neurons. By comparison, although the peak magnitude of LTP is comparable in the young adult and aged hippocampus, the intensity and frequency of stimulation necessary to achieve that response increase with advanced age. Once established, LTP also decays to prepotentiated baseline levels more rapidly in aged subjects than in young animals. This enhanced rate of decay is correlated with the rapid forgetting aged rats exhibit on tests of spatial memory, consistent with the conclusion that a reduced capacity for synaptic enhancement in the hippocampus contributes to this behavioral deficit. Hippocampal LTD is also affected, demonstrating that multiple plasticity mechanisms are altered in relation to variability in the cognitive outcome of aging.

Electrophysiological studies of ongoing neuronal activity in awake-behaving rats provide a direct window on how normal aging influences hippocampal information processing. As described in Chapter 48, a particularly well-characterized phenomenon is that the firing rate of individual pyramidal neurons in the hippocampus increases dramatically as animals navigate through a restricted area within a testing environment. This "place field" activity is largely independent of the specific behaviors that accompany exploration and instead reflects the cognitive demands of testing in relation to sensory cues that define a particular spatial location. Recapitulating a theme from studies of LTP and LTD, many aspects of place field firing are unchanged in the aged hippocampus, including the percentage of neurons that exhibit location-specific firing, and the spatial selectivity of place fields. The overall scope of information encoded, however, is reliably altered (Burke & Barnes, 2006; Wilson et al., 2006). For example, hippocampal pyramidal cells in old rats are abnormally prone to confuse familiar environments that include overlapping stimulus elements, unpredictably engaging the wrong distribution, or "map," of place fields during successive bouts of exploration in a given setting. Hippocampal encoding also becomes increasingly rigid with age such that a relatively narrow subset of available spatial cues comes to control place field activity. These changes are more prevalent in the CA3 field of the hippocampus than in CA1, and among aged individuals that demonstrate impaired spatial learning (Wilson et al., 2006). Research on the structural integrity of the hippocampus, discussed in the following sections, points to a potential link between the behavioral and electrophysiological consequences of aging.

Early Findings Suggested Neuron Loss Is Distributed Diffusely throughout the Aged Brain

One of the most widely held notions about brain aging is that substantial numbers of neurons inevitably die as we grow older. Early studies supported this view, suggesting that neuron death occurs throughout life, with the cumulative loss exceeding 50% in many neocortical areas by age 95 (for an historical review, see Brody, 1970). Although the magnitude of effect appeared to vary, significant

age-related neuron loss was reported for all regions examined, including both primary sensory and association areas of the cortex. On the basis of these early observations it seemed reasonable to suppose that widely distributed neuron death might account for many of the cognitive deficits associated with normal aging. A significant concern, however, was that the finding of widespread neuronal loss might be attributable to preclinical dementing disease in an unknown proportion of the aged individuals who were examined. This confound is overcome in studies using animal models of cognitive aging because rats and monkeys do not develop dementia spontaneously. In addition, as described next, improved quantitative tools for examining neuron loss in the aged brain are now widely available.

Memory Impairment during Normal Aging Does Not Require Substantial Hippocampal Neuron Loss

Methodological developments in quantifying cell number have prompted significant revision in traditional views on age-related neuron loss (Morrison & Hof, 1997). Before the 1990s, most investigators focused on cell density, defined as the number of neurons present in a fixed area or volume of tissue. A significant limitation, however, is that density can vary widely in the absence of any actual difference in cell number. Assume, for example, that total neuron number is identical in two brains, but that the overall size of the brains differs due to normal biological variability among individuals, gliosis, white matter abnormalities, or other neuropil alterations. Under these conditions, neuron density will be lower in the larger brain, simply as a consequence of the cells being distributed in a larger volume. In this way, real or processing-related volumetric differences between young and aged brains could substantially influence cell density in the absence of any actual age-dependent difference in neuron number.

The field of stereology has provided standardized tools for quantifying total neuron number in any defined brain region, yielding an unequivocal measure for examining potential neuron loss during normal aging (see Box 43.1). In contrast to early studies measuring cell density, the conclusion from investigations using modern stereological methods is that the total number of principal neurons (i.e., granule cells of the dentate gyrus and pyramidal neurons in the CA3 and CA1 fields) is preserved in the aged hippocampus. Substantial sparing is also seen in adjacent parahippocampal cortical regions that critically participate in memory. Whereas similar results have been observed in humans, data from animal models are particularly illuminating in demonstrating that hippocampal neuron number remains stable even among aged individuals with pronounced learning and memory deficits. As illustrated in Figure 43.5, for example, stereological analysis reveals that the total number of neurons comprising the hippocampal region is comparable in young adult subjects and aged rats, regardless of their capacity for spatial learning and memory.

The dentate gyrus of the hippocampus is one of only a few niches of neural stem cells capable of generating new neurons in the adult brain (see Box 14.1). Whereas diffuse neuron death is no longer considered a prominent feature of normal aging, considerable interest centers on the idea that a decline in hippocampal neurogenesis contributes to cognitive aging, and that interventions designed to enhance neurogenesis might protect against age-related memory decline, or even rescue function in Alzheimer's disease. Although neurogenesis rates decrease precipitously in the adult rat hippocampus, this effect emerges much earlier in the lifespan than age-related deficits in hippocampal memory, and it fails to predict individual variability in the severity of impairment (reviewed in Bizon and Gallagher, 2005). Considerable additional study will be needed in order to fully exploit neural stem cell biology in support of healthy cognitive aging.

Hippocampal Connectivity Is Compromised during Normal Aging

Substantial research has focused on changes in synaptic connectivity and other markers of functional deterioration as potential contributors to normal cognitive aging. The entorhinal cortex originates the major source of cortical input to the hippocampus, giving rise to the so-called perforant path, which synapses on the distal dendrites of dentate gyrus granule cells in outer portions of the molecular layer. The same entorhinal cortex neurons also innervate other fields of the hippocampus, including the most distal aspects of CA3 pyramidal cell dendrites. Multiple lines of investigation suggest that the integrity of these inputs is compromised with age. Stereological quantification of electron microscope images, for example, has documented that aging is associated with significant loss among a morphologically distinct subset of synapses in the outer molecular layer of the dentate gyrus (i.e., the zone innervated by the entorhinal cortex) (Geinisman, de Toledo-Morrell, Morrell, Persina, & Rossi, 1992), together with alterations of synaptic structure (e.g., descreased postsynaptic density length) in other hippocampal subfields. Input originating in the entorhinal cortex conveys much of the neocortically derived information that the hippocampus processes in support of normal memory, and accordingly, it is reasonable to suppose that the disruption of entorhinal–hippocampus connectivity might contribute to the cognitive outcome of aging. Consistent with this proposal, the magnitude of morphological alterations observed in the hippocampal termination zones of the entorhinal cortex is greatest among aged animals with documented deficits on tasks that require the hippocampus and among older rats that exhibit abnormalities in various physiological measures of hippocampal plasticity.

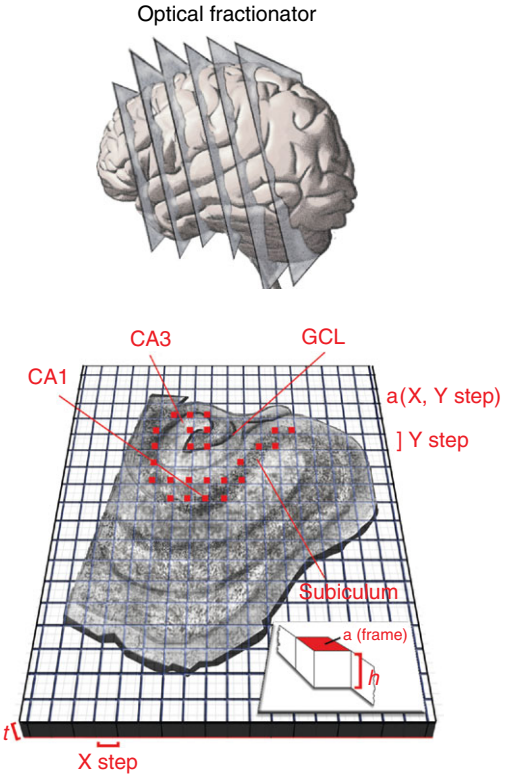
BOX 43.1

THE OPTICAL FRACTIONATOR STEREOLOGICAL METHOD

The figure shows the key features of a stereological technique designed to provide accurate and efficient estimates of total neuron number in a brain region of interest. The hippocampal formation is used as an example. The method consists of counting the number of neurons in a known and representative fraction of a neuroanatomically defined structure in such a way that each cell has an equal probability of being counted. The sum of the neurons counted, multiplied by the reciprocal of the fraction of the structure that was sampled, provides an estimate of total neuron number.

Serial histological sections are prepared through the rostrocaudal extent of the hippocampus and are stained by routine methods for visualizing neurons microscopically. An evenly spaced series of the sections is then chosen for analysis (positions represented schematically in top panel). This first level of sampling, the "section fraction," therefore comprises the fraction of the total number of sections examined. For example, if every tenth section through the hippocampus is analyzed, the section fraction equals 1/10. The appropriate sections are then surveyed according to a systematic sampling scheme, typically carried out using a microscope with a motorized, computer-controlled stage. The lower part of the figure illustrates this design in which the microscope stage is moved in even X and Y intervals, and neurons are counted within the areas defined by the small red squares ("a (frame)" in the inset). The second level of the fractionator sampling scheme is therefore the "area fraction," or the fraction of the XY step from which the cell counts are derived.

The last level of sampling is counting cells only within a known fraction (h) of the total section thickness (t), avoiding a variety of known errors introduced by including the cut surfaces of the histological preparations in the analysis. This is accomplished using a high-magnification microscope objective (usually 100X) with a shallow focal depth. In the illustration provided, the "thickness fraction" is defined as h/t. Neurons are counted as they first come into focus, according to an unbiased counting rule, called the "optical



disector," that eliminates the possibility of counting a given cell more than once.

Finally, total neuron number in the region of interest (N) is estimated as the sum of the neurons counted (sumQ–), multiplied by the reciprocal of the three sampling fractions; the "section fraction," "area fraction," and the "thickness fraction." For the present example, the total estimated neuron number is given by

$$N = (\text{sumQ}-) \times (10/1) \times (\text{XY step area/a(frame)}) \times (t/h).$$

Original illustration design by L. E. Mekiou and P. R. Rapp.

Peter R. Rapp

Executive Function Mediated by the Prefrontal Cortex Declines during Normal Aging

Normal aging is sometimes characterized as a generalized process of nonspecific deterioration across many neural systems, implying that age-related cognitive decline might be the cumulative effect of subtle alterations distributed diffusely throughout the brain. Little empirical evidence actually supports this notion, however, and as we have seen in the case of hippocampal circuitry, the current perspective is that the neurobiological consequences of aging exhibit considerable selectivity. At the same time, it is important to recognize that the effects of aging on memory do not occur in isolation and that cognitive capacities mediated by other neural systems are also vulnerable.

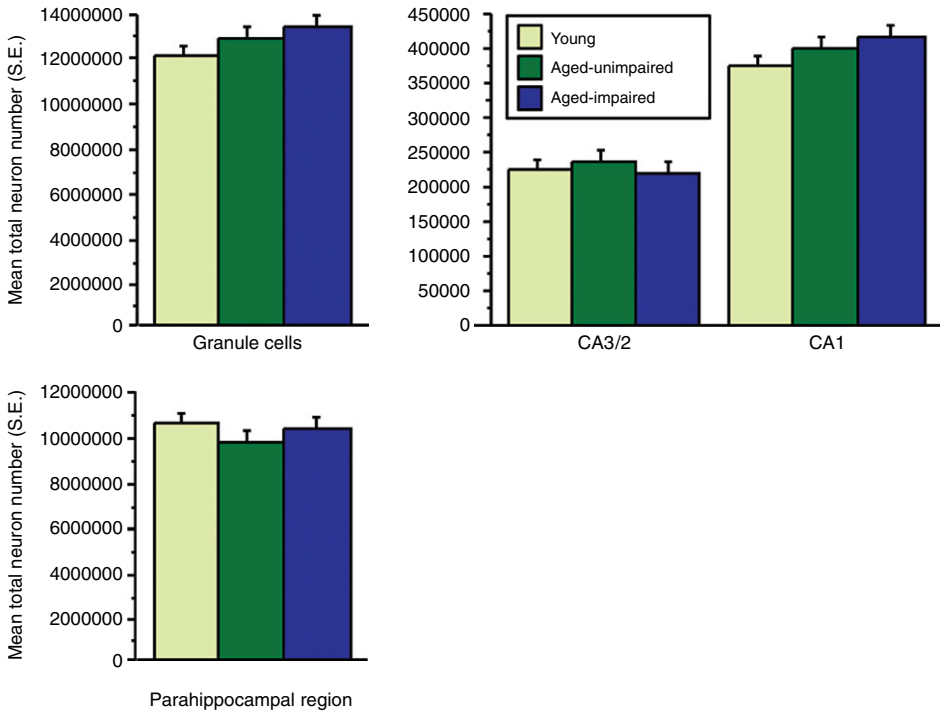


FIGURE 43.5 Estimated total neuron number (plus S.E.) in principal cell layers of the hippocampus and parahippocampal cortices (i.e., entorhinal, perirhinal, and postrhinal) for behaviorally characterized young and aged rats. All values are unilateral. Half of the aged rats exhibited substantial spatial learning deficits in the Morris water maze (aged impaired); the other half performed within the range of learning scores for the young group (aged unimpaired). Neuron number is stable with age and across a broad range of learning capacities. *Adapted with permission from Rapp and Gallagher (1996) and Rapp, Deroche, Mao, and Burwell (2002).*

As discussed in Chapter 50, the functionally and neuroanatomically heterogeneous areas comprising the prefrontal cortex support a constellation of capacities referred to as "executive function." Whereas damage to these regions does not produce the frank amnesia that follows medial temporal lobe lesions, significant deficits are observed in a variety of processing functions that influence memory, including the strategic use and manipulation of remembered information, the ability to recall the source from which information was acquired, and the order in which it was presented. It is therefore of interest that a qualitatively similar profile of deficits emerges during the course of normal aging (reviewed in Fletcher and Rapp, 2012). Older people, for example, often experience difficulty recalling the original source of remembered information, even when recollection of the target item itself is successful. These failures in "source memory" predict performance on other tests of frontal lobe function, including memory for temporal order, suggesting they share a common neurobiological basis.

Neuropsychological studies in animal models reinforce the conclusion that prefrontal cortex dysfunction contributes to certain features of cognitive aging. One of the best-documented deficits associated with aging in monkeys, for example, is poor performance under testing conditions that emphasize the spatial and temporal, or the "where and when," components of memory (Fletcher & Rapp, 2012). In the classic delayed response task, subjects observe while a food treat is hidden in one of two locations that are then covered with identical plaques (see Fig. 50.5). Because each reward location is baited frequently across trials within a single test session, there is substantial opportunity for interference and accurate performance depends on the ability to keep in mind the spatial location that was baited most recently. Damage involving the dorsolateral prefrontal cortex in young monkeys produces impairment on the delayed response procedure at retention intervals of just a few seconds or more. Deficits on this task following lesions of the hippocampal system, in contrast, are selectively observed at longer delays. Numerous studies since the late 1970s have documented that aged monkeys exhibit a profile of delayed response impairment qualitatively similar to the effects of prefrontal cortex lesions in young animals. Additional signs of prefrontal decline are also found in aged monkeys, including marked perseveration and behavioral rigidity when subjects are challenged with shifting task contingencies. For example, whereas young adults quickly cease responding to a previously positive stimulus when it is no longer rewarded, aged monkeys are slower to make this behavioral adjustment.

Are the effects of normal aging on prefrontal cortex capacities coupled to other signatures of cognitive decline, consistent with the view of aging as a process of generalized deterioration? Available evidence suggests that this is not the case. For example, age-related deficits on tests of prefrontal cortex function tend to emerge earlier during the course of aging than alterations in declarative memory supported by the medial temporal lobe system. In addition, when impairments indicative

of both frontal and medial temporal lobe dysfunction cooccur in the same subjects, these effects are statistically unrelated to each other (reviewed in Fletcher and Rapp, 2012). Thus, different aspects of cognitive function can decline independently during aging and these effects are not the consequence of a global, brain-wide degenerative process.

Neurobiology of Aging in the Prefrontal Cortex

The basis of age-related decline in capacities supported by the prefrontal cortex is the subject of intense investigation. Similar to the hippocampal system, neuron death is not a prominent feature of normal aging in many frontal lobe areas that mediate information processing functions known to decline with age. Other structural alterations have been documented, however, including a substantial decline in the density of prefrontal cortex synapses, marked changes in dendritic architecture, and a variety of abnormalities in the myelination of axons (Peters, Morrison, Rosene, & Hyman, 1998). These parameters are correlated with behavioral signatures of cognitive aging in monkeys, raising the possibility that, like the hippocampal system, changes in the functional organization of critical prefrontal networks may play a significant role. Consistent with this possibility, recent findings document that delayed response impairment in the aged monkey is associated with the blunting of memory-related neuronal activity in the dorsolateral prefrontal cortex, and that normal task dependent firing is at least partly restored by pharmacological treatments that benefit behavioral performance (Wang et al., 2011).

Results from *in vivo* imaging studies in humans complement the animal model data and further suggest that the functional organization of the aged brain is more dynamic than previously presumed. The traditional understanding has been that successful cognitive aging reflects the persistence of a youthful condition, dependent on the same mechanisms that support normal function in younger individuals. Rather than following a simple pattern of regional decline or maintenance, however, growing evidence indicates that aging is accompanied by reliable change in the distribution of activations across brain areas important for normal cognitive function, including the prefrontal cortex and medial temporal lobe. Indeed shifts in the balance of regional activation are observed even when memory is tested under conditions where aged individuals score as accurately as young adults. The implication of these findings is that the overall network engaged during testing is significantly altered and that successful cognitive outcomes in normal aging may rely, in part, on adaptive functional reorganization (reviewed in Hedden and Gabrieli, 2005). Intervention aimed at harnessing this dynamic capacity in support of optimally healthy cognitive aging is an area of active investigation.

Neurochemically Specific Subcortical Systems Are Susceptible to Aging and Positioned to Broadly Influence Behavior

Current evidence challenges the view that normal aging is associated with generalized deterioration distributed diffusely throughout the brain. Aging does, however, affect the integrity of neurochemically specific classes of cells whose ascending projections influence widespread brain areas. Cholinergic cell groups in the basal forebrain have been studied particularly intensively because this system exhibits pronounced degeneration in Alzheimer's disease. Less severe effects are also observed during normal aging, involving acetylcholine-containing neurons that project to the hippocampus, amygdala, and neocortex. Cholinergic cell loss might therefore disrupt the information processing functions of these target regions, and consistent with this possibility, reliable correlations have been reported between the magnitude of cholinergic deficits and behavioral impairment in aged individuals. Cholinergic abnormalities alone, however, fail to account for the full profile of cognitive deficits observed during aging since neurotoxin lesions that selectively destroy these neurons fail to reproduce key features of the behavioral decline observed in older individuals. One possibility is that age-related cholinergic impairment comprises an important component of a broader constellation of alterations that leads to cognitive dysfunction. Other neurochemical systems that might contribute include dopamine-containing circuitry that originates in midbrain cell groups and noradrenergic inputs that arise from the locus coeruleus. These systems innervate a variety of more anterior brain regions, and consistent with this broad distribution, neurochemical alterations have been linked to a diverse array of behavioral deficits ranging from age-related cognitive impairment to disturbances in motor function. These findings encourage the view that appropriately designed pharmacological interventions might have beneficial effects on a similarly broad profile of outcome measures of aging.

Summary

This section considered the cognitive and neurobiological effects of normal aging in humans and animal models. Key themes include the observation that the impact of aging on memory varies widely from one individual to the next. Whereas neuron death was once presumed to be the proximal cause of many cognitive deficits associated with aging, including learning and memory impairment, it is now clear that such decline does not require marked neuron loss in the hippocampus and related cortical areas. Instead, subtle alterations in connectivity, and downstream changes in cellular function, are more likely causative factors. Fueled by advanced *in vivo* brain imaging studies, a comprehensive, neural systems account of normal cognitive aging is beginning to emerge.

PATHOLOGICAL PROCESSES IN COGNITIVE DEVELOPMENT AND AGING

Developmental Psychopathologies

Research in this field focuses on elucidating the interplay among the biological, psychological, and social-contextual aspects of normal and abnormal development across the life span. Recent advances in neuroimaging and gene linkage technology have vastly expanded research possibilities for noninvasive clinical studies. Even with a new arsenal of research technologies, however, it has not been possible to isolate and define characteristics of each developmental disability. Such efforts are limited by several factors:

1. Disagreement regarding the criteria for diagnosis. Because different researchers often use differing behavioral criteria, cross-study comparisons frequently show inconsistencies in results.
2. Heterogeneity of disorders. Even when consistent behavioral criteria are carefully applied, different underlying etiologies can result in the same behavioral profile.
3. Difficulties inherent in the study of children. Although modern brain imaging techniques such as magnetic resonance imaging are considered noninvasive, they are stressful and time-consuming, and parents will not always allow their affected children to participate in such studies. Although studies can be conducted using adults who have been affected since childhood, data obtained from such studies may not accurately reflect anomalies that characterize the early disruption of brain development because brains change over time. Moreover, retrospective diagnoses of childhood disorders frequently rely on memory (e.g., a patient is asked, "Did you have difficulty reading as a young child?") and thus can be unreliable.

Despite these challenges, ongoing research has uncovered important neurological and genetic features that seem to be associated with specific developmental disorders (see Box 43.2). The following sections outline what is known about how specific injuries or anomalies in development of the brain may result in specific patterns of cognitive and behavioral impairment. Two examples of developmental psychopathology are described: autism and schizophrenia.

Autism

Autism is a neurodevelopmental syndrome first described by Kanner in 1943, characterized by deficits in multiple domains including social interaction, play, language, and communication, and a restricted range of interests and activities, including repetitive and stereotyopic movements and persistent preoccupation that begin in infancy and become more apparent by the end of the third year of life. Although autism was once considered a rare disorder, recent surveys report a prevalence of 40 to 50 per 10,000 births. Parents often first become concerned because their child fails to use words to communicate, even though he/she recites passages from videotapes or says the alphabet. Like other disorders of abnormal cognitive development, autism is a heterogeneous condition that can range widely in severity.

Characteristics of Autism

A multitude of symptoms occur in autism, including social isolation (the autistic child largely ignores other people, shows little attachment to parents or other relatives, and retreats into a world of his or her own); stereotyped behaviors (the autistic child rocks back and forth, stares at neutral stimuli, rotates an object, or engages in other repetitive behaviors for long, uninterrupted periods); resistance to change in routine; abnormal responses to sensory stimuli (the autistic child may ignore visual stimuli and sounds, especially speech sounds, sometimes to such an extent that it might be thought the child is deaf); inappropriate emotional expressions (the autistic child has sudden bouts of fear and crying for no obvious reason; at other times the child displays utter fearlessness and unprovoked laughter); and poor use of speech. Individuals with autism can also exhibit symptoms that fall along a continuum of disorders. For example, many have mental retardation and show delays in the development of language. Others display neither language delay nor mental retardation, although they are clearly socially inept. Finally, autism is also associated with idiosyncratic interests and a restricted and repetitive behavioral repertoire. A growing number of studies over the last two decades have provided evidence that autism results from early onset brain dysfunction.

Neuropathological Alterations in Autism

Several studies of autistic individuals provide evidence for megalencephaly (whole brain enlargement) and for increased head circumference. The nature of this enlargement has yet to be studied carefully, but it is specifically observed in the occipital, parietal, and temporal lobes. Postmortem analyses of the brains of autistic individuals have revealed increased cortical thickness and abnormalities in cortical morphology. Furthermore, abnormalities in the size and density of neurons in the medial temporal lobe and limbic system, including the amygdala, hippocampus, and entorhinal cortex, septal nuclei, and cingulate cortex, have also been documented. These neuropathological abnormalities make sense with respect to certain major symptoms of autism because the limbic system is known to play a central role in coordinating social–emotional functioning. An fMRI study of autism has documented dysfunction localized to the ventral temporal lobe using a face recognition task (Baron-Cohen et al., 1999). In this study,

BOX 43.2

DYSLEXIA

Reading skills occupy a uniquely important position in overall cognitive development, strongly dependent on antecedent skill acquisition and critically supportive of educational success. After accounting for the effects of inadequate instruction, the most common cause of severe reading problems in childhood is *developmental dyslexia*, a disorder typically characterized by impairment in applying the sound–correspondence rules necessary to decode print (see Chapter 49). Although reading impairment is common, with an incidence of 5 to 10% among schoolage children, our understanding of its biological roots is relatively recent.

Neurobiological studies of dyslexia are challenging, as the behavioral manifestations of this disorder are complex. Although defined by poor reading performance, dyslexia involves additional deficits, including poor phonological processing, poor verbal working memory, and slow naming ability. Dyslexics frequently show subtle deficits in motor control and early sensory processing, most notably in the visual system. Imaging studies have revealed that dyslexic brain structure differs from that seen in individuals with normal reading skills. Micro and macroscopic structural abnormalities have been detected in perisylvian language regions (temporal and parietal banks of the sylvian fissure as well as the insula), visual system structures, the thalamus, and the corpus callosum. The planum temporale is an expanse of neocortex on the temporal bank of the sylvian fissure, its anterior border defined by Heschl's gyrus. From the time of birth the left planum temporale is larger than the homologous region in the right hemisphere in 70 to 80% of individuals. Reduction or reversal of this leftward asymmetry has been documented in individuals with developmental dyslexia. The functional concomitants of these anatomical variations have been explored in functional imaging studies that have revealed differences in patterns of task-related activity between dyslexics and controls during the performance of reading and phonological decoding tasks. Interestingly, the cortical areas identified are similar in English, French, and Italian dyslexics, supporting a biological origin for this reading abnormality, independent of cultural upbringing. Similarly, candidate genes for dyslexia have been identified and replicated in different countries, suggesting the multigenetic contribution to reading disability.

Besides the problems related to reading, dyslexics also exhibit subtle abnormalities in visual processing. Evidence for selective involvement of magnocellular pathways of the visual system (see Chapter 26) has accumulated, beginning with the demonstration that cell bodies in the magnocellular layers of the lateral geniculate nucleus are smaller in dyslexics compared to a control group. The fact that both contrast sensitivity and visual persistence are abnormal in reading disabled children is consistent with the notion that these children have disturbances in the magnocellular system, which is specialized for temporally demanding visual processing. These behavioral results are supported by functional imaging studies demonstrating less task-related activity in extrastriate and parietal cortex during visual motion detection by dyslexics.

Future studies will determine the nature of the neural mechanisms that link deficits in early visual processing with more cognitive skills such as phonological awareness. Also remaining to be clarified are issues concerning which of these deficits is a primary cause of dyslexia as opposed to a consequence of having partially compensated for a reading disorder. Such information will aid in identifying better avenues for reading remediation. To date, the best methods for improving the reading abilities of dyslexics focus on structured teaching of the code that allows individuals to sound out words (phonological awareness) in the context of fluency and comprehension exercises.

Guinevere F. Eden

high-functioning individuals with autism and matched controls were presented with a series of photographs of eyes while their brains were being scanned to assess brain regions activated by two different tasks. In the first task (experimental condition), each photograph was presented simultaneously with words (either unconcerned or concerned, or sympathetic or unsympathetic). Participants were asked to indicate by pressing a button which word best describes what the person in the photograph is feeling or thinking. In the second task (control condition), each photograph was presented with the words "male" or "female" and the subjects pressed a button to indicate gender. In both tasks, the performance of autistic individuals and controls was above chance. However, the controls were more accurate in the two tasks than the autistic individuals. To investigate the brain regions activated by the experimental task, brain maps obtained during the control task were subtracted from brain maps obtained during the experimental task. In the control group, the experimental task activated two main regions of the brain: the frontotemporal neocortical areas, including the left dorsolateral prefrontal cortex and supplementary motor areas, bilateral temporoparietal regions, and subcortical structures, including the left amygdala and parahippocampal gyrus.

The autism group activated the frontal components but did not activate the amygdala. The significance of these data is that one critical role of the amygdala is to identify the mental states or emotions of other individuals and that this structure appears to be dysfunctional in persons with autism (see Pelphrey and Carter (2008), for additional information on the neuroanatomical substrates of social cognition dysfunction in autism).

Significant atrophy of the vermal lobules in the cerebellum has also been reported in autistic individuals, although this effect does not appear specific to autism and is found widely in neurodevelopmental disorders. By comparison, there is now extensive evidence for the involvement of higher association areas in the pathophysiology of autism, particularly the frontal and parietal cortices, as well as significant remodeling of cortical interconnectivity (Raznahan, Toro et al., 2010). Cytoarchitectural abnormalities that have been described in the cortex of patients with autism include a thickened cortex, high neuronal density, minicolumn alterations, the presence of neurons in the molecular layer, irregular laminar patterns, and poor gray-white matter boundaries.

Autism Is a Complex Genetic Disorder

The most compelling evidence for a genetic contribution to autism comes from studies of monozygotic (identical) and dizygotic (fraternal) twins. When one member of a pair of monozygotic twins is autistic, the other twin has more than a 50% probability of being autistic. However, if one member of a pair of dizygotic twins is autistic, the other twin has only a 3% probability of developing autism. Using conventional nomenclature, we say that monozygotic twins have a 50% or higher concordance for autism and that dizygotic twins have a 3% concordance. The relative risk factor for siblings is one of the highest for a complex genetic disorder. The most significant genetic finding relevant to autism is the recent identification of genes responsible for Rett's syndrome and fragile X syndrome, two conditions associated with autistic features. Rett's syndrome is caused by mutations in the methyl-CpG binding protein 2 (MECP2) gene. Decreased expression of this gene leads to a failure to suppress expression of genes regulated by methylation. Fragile X syndrome is a disorder caused by mutation of the gene FMR1 (see Box 43.3). A better understanding of how reduced expression of the MECP2 and FMR1 genes leads to mental retardation as well as social and language disorders will likely provide a better understanding of the autistic disorder. Over the last several decades, a growing list of rare mutations, such as *NGLN4X, SHANK 3, NRXN1,* have been described in individuals with autism (State, 2010). However, few of these genes are specific to autism and they instead variably contribute to genetic risk for a variety of neuropsychiatric disorders, including schizophrenia and attention-deficit

BOX 43.3

TRIPLET REPEAT DISORDERS

Not all neurogenetic diseases follow simple inheritance patterns; not all developmental disorders reveal themselves immediately; and diseases with similar genetic etiologies can have widely different manifestations. *Trinucleotide repeat mutations*, or *dynamic mutations*, were discovered only in the early 1990s but are responsible for a number of important neurodegenerative disorders, such as Huntington Disease (see Box 31.4), Fragile X mental retardation, and myotonic dystrophy. To date, 14 neurological disorders have been found to result from such a mutational mechanism, and the list will probably continue to grow.

Triplet repeat mutations are unstable and prone to expand as alleles are passed from one generation to the next, although contractions can occur. Repeats at different loci differ markedly in their rates and extents of expansion, and the instability of repeats at each locus often depends upon the sex of the transmitting parent.

Fragile X syndrome, an X-linked dominant mental retardation syndrome, was the first disease shown to result from a triplet repeat mutation. There is a CGG repeat expansion at a locus (Xq27.3) giving rise to a folate-sensitive fragile site (FRAXA). The CGG repeat ranges in length in the general population from 5 to 50 triplets. Affected individuals carry more than 230 repeats, up to several thousand triplets. CGG tracts between 45 and 200 repeats in length are termed "premutations"; they are not long enough to result in mental retardation, yet are prerequisite for further expansion to the full mutation. Escalating risk of affected offspring is associated with increasing size of premutations in female carriers. Curiously, offspring of male carriers of premutations are not at risk for Fragile X syndrome, as these alleles have not been observed to expand to full mutations upon male transmission. The normal function of the affected protein FMR1 remains to be fully determined, but it appears to play a role in RNA metabolism, likely in the areas of transport, stability, and/or translation.

Myotonic dystrophy (DM) remains a particularly enigmatic triplet repeat disorder affecting multiple systems, causing myotonia, muscle weakness, cardiac conduction defects, diabetes, cataracts, premature balding, and sometimes dysmorphic features and mild mental retardation. There is a tendency for these features to appear in more

BOX 43.3 (*cont'd*)

severe form in subsequent generations of a family. In DM there is a CTG repeat in the 3′ untranslated region of the myotonin protein kinase gene (DMPK). This CTG tract is small in the general population (5–37 triplets) but can expand dramatically to thousands of repeats. Slightly increased lengths (70–100) can result in mild effects, such as late-onset cataracts, and the classic symptoms of early adulthood myotonia are found in individuals with hundreds of repeats. Very large expansions can cause severe disease in newborns.

The mechanism(s) by which the mutation leads to the various clinical symptoms remains uncertain. A variety of effects have been described and the repeat expansions clearly affect both the DMPK gene and nearby genes. It is also likely that the expanded CTG repeat carrying mRNAs can act to diminish the levels of nuclear proteins that bind at these sequences, resulting in misregulation of genes not linked to DMPK. Mutations in equivalent mouse genes have not provided a complete model of the human disease. The phenotype in DM results from multiple effects of the repeat expansion, rather than a simple loss or gain of function mechanism.

Friedreich's ataxia (FA) causes both limb and gait ataxia, loss of position and vibration sense, decreased tendon reflexes, cardiomyopathy, diabetes, and, in some patients, optic atrophy and deafness. FA is a recessive disorder and the only triplet repeat disorder known to involve a GAA sequence. Carriers of FA are rather common (1/85), and, consequently, so is the GAA expansion. Members of the general population carry between 7 and 34 repeats, while pathogenic alleles can number in the hundreds. The affected frataxin gene product appears to be involved in mitochondrial iron homeostasis, and its absence affects postmitotic cells rich in mitochondria. The mechanism by which the GAA repeat affects gene function appears to be inhibition of primary RNA transcripts.

A second class of triplet repeat disorders includes those caused by expansion of translated CAG repeats which encode a polyglutamine tract in each of the respective proteins. These "polyglutamine" disorders share many features, suggesting that a common pathogenetic mechanism is at play in spite of the fact that the mutated genes share no homology outside of the CAG repeats. They are progressive neurologic diseases with onset of symptoms in young to midadulthood with only a specific group of neurons vulnerable in each disease. Symptoms develop when the number of uninterrupted repeats exceeds approximately 35 glutamines. The phenotype is caused not by loss of function of the relevant protein but rather by a gain of function conferred by the expanded polyglutamine tract.

The *spinocerebellar ataxias (SCA)* are genetically distinct but overlap a great deal in their clinical presentation—between subtypes and even within families affected by the same disease. It is nearly impossible to distinguish one subtype from another on the basis of clinical signs alone. *Spinocerebellar Ataxia Type 1 (SCA1)* was the first of the inherited ataxias to be mapped. Patients with SCA1 suffer from progressive ataxia (loss of balance and coordination), dysarthria, and eventual respiratory failure. SCA1 is characterized pathologically by cerebellar atrophy with severe loss of Purkinje cells and brain stem neurons. Patients with SCA1 have an expanded "perfect" CAG repeat tract encoding ataxin-1, the function of which is largely unknown.

Studies of human patients with deletions of, for example, the HD, or SCA1 genes (and of mice lacking these genes) confirm that disease is not due to loss of function of the respective proteins. More importantly, studies of mouse and fruit fly models that over-express full-length or truncated forms of either ataxin-1, ataxin-3, huntingtin, AR, atrophin-1, or ataxin-7 demonstrate progressive neuronal dysfunction consistent with a gain of function mechanism. Several illuminating findings have emerged from the study of human patients, cell culture systems, and the various animal models. In transgenic mice, expression of truncated polypeptides containing expanded glutamine tracts caused widespread dysfunction that extends beyond the specific groups of neurons affected by the full-length protein. The expanded polyglutamine tracts may cause the respective proteins to misfold or resist degradation. Aberrant protein-protein interactions or alterations in levels of proteins essential for the normal functioning of neurons (such as proteins that regulate Ca^{+2} levels and neurotransmitters) could, in turn, exacerbate neuronal dysfunction and lead to neuronal death.

The finding that chaperone overexpression mitigates the phenotype in one vertebrate and several invertebrate models of polyglutamine disorders, and the identification of several additional modifiers of the neuronal phenotype, provide a platform for identifying key pathways of pathogenesis and potential therapeutic targets.

Adapted by Graham V. Lees with permission from Nelson, D., & Zoghbi, H. (2003). Disorders of Trinucleotide repeat expansions. In Encyclopedia of the human genome. *London: Nature Macmillan.*

Suggested Readings

Cummings, C. J., & Zoghbi, H. Y. (2000). Trinucleotide repeats: Mechanisms and pathophysiology. *Annual Review of Genomics and Human Genetics, 1,* 281–328.

Wells, R. D., & Warren, S. T. (Eds.), (1998). *Genetic instabilities and hereditary neurological diseases*. San Diego: Academic Press.

hyperactivity disorder (ADHD). Although the findings provide tremendous promise, they also highlight the many challenges facing the study of the genetics of autism.

Schizophrenia

Schizophrenia is a catastrophic illness with an onset typically in adolescence or early adulthood. It was identified and defined by Eugen Bleuler in 1911. The combination of significant incapacity, onset early in life, and chronicity of illness makes schizophrenia a particularly tragic disorder, occurring in 0.5 to 1% of the population.

Characteristics of Schizophrenia

The clinical symptoms of schizophrenia are often divided into two broad categories: positive and negative. Positive symptoms include delusions, hallucinations, disorganized speech, and disorganized or bizarre behavior. These symptoms are referred to as positive because they represent distortions or exaggerations of normal cognitive or emotional functions. The negative symptoms of schizophrenia reflect a loss or diminution of normal function. Negative symptoms include alogia (marked poverty of speech or speech devoid of coherent content), affective flattening (diminution in the ability to express emotion), anhedonia (inability to experience pleasure); avolition (inability to initiate or persist in goal-directed behavior), and attentional impairment. Although the positive symptoms of schizophrenia are often colorful and draw attention to the patient's illness, the negative symptoms are an important feature of the disease that impair the person's ability to function normally in daily life.

The profound and pervasive cognitive and emotional disturbances that characterize schizophrenia suggest that it is a serious brain disease affecting multiple functions and systems. Auditory hallucinations and disruptions in linguistic expression suggest involvement of the auditory cortex and perisylvian language regions. Other positive symptoms, such as delusions or disorganized behavior, are more difficult to localize to specific regions and suggest dysfunction distributed across multiple neural systems and circuits. Negative symptoms may be related to the prefrontal cortex, which mediates goal-directed behavior and fluency of thought and speech. The affective flattening and asociality suggest involvement of the limbic areas within the temporal lobe.

The Dopamine Hypothesis of Schizophrenia

A great deal of research since the 1950s has made it clear that schizophrenia is associated with abnormalities in the dopaminergic synapses of the brain. Still, the exact nature of this perturbation remains elusive. The strongest link between dopamine synapses and schizophrenia comes from studies of drugs that alleviate the symptoms of schizophrenia. A large number of antischizophrenic, or neuroleptic, drugs have been found, most of which belong to two chemical families, the phenothiazines, which include chlorpromazine, and the butyropherones, which include haloperidol. All of these drugs have two properties in common: they block postsynaptic dopamine receptors and they inhibit the release of dopamine from the presynaptic neuron. One interpretation of these results is that schizophrenia is due to excess activity at dopamine synapses. Establishing cause-and-effect relationships in human psychiatric disorders, however, is challenging. For example, although a substantial increase in the number of dopaminergic receptors is observed in the schizophrenic brain, because virtually all patients receive neuroleptic medication, it is not known if increased receptor density is a consequence of the disease itself or a response to the extended use of antidopaminergic drugs. An additional complication for the dopamine receptor hypothesis of schizophrenia concerns the time course for the effects of neuroleptic drugs. Although these treatments block dopamine receptors almost at once and reach their full pharmacological effectiveness within a few days, effects on behavior emerge more slowly, over 2 or 3 weeks. Clearly the symptomatic relief achieved with neuroleptic intervention involves mechanisms in addition to the blockade of dopamine receptors.

One possibility is that the prolonged use of neuroleptic drugs decreases the number of spontaneously active dopamine neurons in what is known as the mesolimbic system, a set of neurons that projects from the midbrain tegmentum to the limbic system. Another possibility is that the underlying problem in schizophrenia is not an excess of dopamine activity at all, but a deficit of glutamate activity. Glutamate is the predominant excitatory amino acid released by neurons in the cerebral cortex that project widely throughout the limbic system, and dopamine synapses are known to inhibit glutamate release in these target regions. If glutamate release is reduced in the schizophrenic brain, one way to correct this deficiency would be to block activity at dopaminergic receptors, relieving glutamate synapses from inhibition.

Neuropathological Alterations in Schizophrenia

The most consistently described structural alterations in schizophrenia are ventricular enlargement and decreases in the volume of temporal lobe structures, including the hippocampus, amygdala, and entorhinal cortex. Cytoarchitecture and cell densities are also reportedly altered in limbic structures: reduced cell number and size have been observed in the hippocampus, parahippocampal gyrus, and entorhinal cortex, along with disturbed cytoarchitecture involving cellular disarray in the hippocampus and the entorhinal and cingulate cortices. This distribution of structural abnormalities may help account for the symptoms of schizophrenia in that these limbic structures

link neocortical association areas with the septum–hypothalamic complex and are therefore key in sensory information processing and sensory gating. Defects in this neural circuit could lead to a dissociation between neocortically mediated cognitive processing and limbic–hypothalamic emotional reactions and to a disturbed emotional experience of external sensory perceptions.

The structural alterations reported in schizophrenia are often subtle and certainly heterogeneous, contributing minimally to the elucidation of the disease if symptoms are to be attributed to defects in individual brain regions. Alternatively, it is currently hypothesized that the variety and severity of symptoms observed in schizophrenia can be explained by defects in several brain regions affecting the neural circuits that mediate higher cognitive functions. Such neural circuitry would include neocortical association areas, the limbic system and midline thalamic structures, with possible involvement of the basal ganglia.

The Neurodevelopmental Hypothesis of Schizophrenia

Several clinical observations point to neurodevelopmental abnormalities in schizophrenia: onset in adolescence in most patients and earlier onset among males, pronounced premorbid neurodevelopmental abnormalities, such as asociality and "soft" neurological signs, impaired cognitive and neuromotor functioning, minor physical anomalies, presence of structural abnormalities at the onset of the illness, which may predate the illness, and the absence of neurodegenerative processes.

Postmortem morphological observations in the brains of schizophrenic patients suggest that multiple aspects of neural development may be disrupted. The spatial disorganization of neurons found in the hippocampus and the inappropriate laminar location of cells in the entorhinal cortex, for example, suggest a potential defect in neuronal migration. Other evidence indicates that the organization of synaptic connectivity is affected. The functional circuitry of the frontal cortex undergoes significant remodeling until adolescence, mediated by an NMDA receptor-dependent mechanism. The number of these receptors is reduced in the brain of schizophrenics. Still other abnormalities involve biochemical alterations of membrane phospholipids in the dorsolateral prefrontal region. These changes include a decrease in phosphomonoesters and an increase in phosphodiesters. Such biochemical changes in membrane phospholipids are known to occur during normal synaptic pruning in adolescence, but appear to be exaggerated in the dorsolateral prefrontal cortex of schizophrenics. These enhanced biochemical changes may lead to greater synaptic pruning in schizophrenics than in normal individuals and contribute to the reduction in prefrontal neuropil volume that accompanies the disease.

The precise timing of neurodevelopmental abnormalities in schizophrenia remains controversial. Some implicate a fixed neural lesion from early life interacting with normal neurodevelopmental events that take place much later. Others posit that schizophrenia is caused by a deviation from normal brain maturational processes during late adolescence, involving large-scale synaptic elimination or pruning in brain regions critical for cognitive development. Still others propose that early brain pathology acts as a risk factor rather than a sufficient cause so that its effects can only be understood in the light of the individual's later exposure to other risk and protective factors.

Genetics Play a Role in Schizophrenia

As for autism, a substantial body of evidence indicates that schizophrenia is associated with both genetic and environmental factors. Findings pointing to a significant genetic contribution come from family, twin, and adoption studies. Evidence for environmental factors derives from studies of prenatal intrauterine environment, neonatal obstetric complications, and postnatal brain insults. Currently, it remains an open question whether these environmental factors play a necessary role in the neurodevelopmental abnormalities responsible for psychotic brain disorders or whether their role is additive or interactive with genetic contributions.

People with schizophrenia are more likely than others to have schizophrenic relatives. The most compelling evidence is the 50% concordance rate for monozygotic twins relative to 15% concordance for dizygotic twins. Neuroimaging studies of discordant monozygotic twins consistently indicate more prominent structural brain abnormalities in the ill twin. More importantly, a neuroimaging study demonstrated that schizophrenia is associated with dysfunction of a prefrontal-limbic network (Weinberger, Berman, Suddath, & Torrey, 1992). In this study, monozygotic twins discordant for schizophrenia underwent PET blood flow scans while they performed a task known to activate the dorsolateral prefrontal cortex in normal subjects. This assessment, known as the Wisconsin Card Sorting Task (WCST), measures abstract problem solving requiring attention and working memory (see Fig. 50.4). The participant sees four cards bearing designs that differ in color, form, and number of elements. The subject's task is to sort the stack of cards into piles in front of the stimulus cards. The subject is told whether the sorting choice is correct or incorrect. The test progresses such that the correct solution is initially stimulus color. Once mastered, the correct sorting strategy is switched without warning to form, requiring subjects to inhibit classifying cards on the basis of color. In a final phase the correct solution again changes unexpectedly, now to the number of elements. Shifting response strategies is difficult for patients with frontal lobe dysfunction. Among affected

schizophrenic twins, prefrontal activation correlates strongly with both left and right hippocampal volumes; a relationship that is not observed in their unaffected twins. In addition, functional connectivity between the hippocampus and prefrontal cortex as measured by fMRI is weaker in schizophrenic patients than controls. These data suggest that schizophrenia may result from dysfunction distributed across a neural network involving prefrontal cortical areas and limbic structures.

Molecular studies demonstrate that schizophrenia is a complex genetic disorder, implicating several possible susceptibility loci, including regions on chromosomes 1q, 6p, 8p, 13q, and 22q. Of genes mapped, the 22q11 gene [a common functional polymorphism of catechol-*O*-methyltransferase (COMT), which is a methylation enzyme that metabolizes released dopamine] has been a popular candidate because of the long-hypothesized role of dopamine in schizophrenia. Recent molecular imaging studies in humans have also established that schizophrenia is associated with alterations in the dopamine system and these neurotransmitter abnormalities are shared by individuals at genetic and clinical risk for developing the illness.

Sex Hormones May Influence Neurodevelopmental Disorders

There are gender differences in the occurrence of neurodevelopmental disorders. The gender ratio of incidence is skewed toward boys for a variety of developmental disabilities, including severe mental retardation (1.3 boys/1 girl), speech and language disorders (2.6/1), learning difficulties (2.2/1), dyslexia (4.3/1), schizophrenia (1.9/1), and autism (4/1). Why? Although the existence of sex differences in the area of language and learning disorders has been contested on the grounds that it is a result of teacher and clinician bias in identifying these disorders, research supports the existence of a significant gender difference in the incidence of these disorders. Moreover, gender differences are also seen in clear-cut phenomena such as complications of pregnancy and birth, which are more common in male infants, and these differences cannot be explained as a reflection of investigator or clinician bias. Several theories attempt to explain these gender differences.

One theory posits that male but not female fetuses invoke a form of "antigenic" response from the mother during pregnancy, being recognized by the immune system as "foreign." The hostile environment thus created for the male fetus *in utero* would explain not only a higher incidence of developmental disorders among boys, but also the finding that later-born sons are increasingly likely to be affected (consistent with increased immune responsiveness on repeated exposure).

Another theory (termed the Geschwind theory, after the neurologist Norman Geschwind) proposed that the uneven gender ratios in developmental disorders are associated with exposure to some "male factor" (possibly androgen) during the last trimester of fetal development. This factor acts to slow cortical maturation in male fetuses, particularly in the left hemisphere, rendering them more susceptible to perturbation during the normal course of development. This exposure to a male hormone would explain the higher incidence of developmental disorders among boys. Conversely, faster central neurons system development in female fetuses would enable them to better withstand insult during late pregnancy and birth. This idea is supported by evidence that female infants appear to show better cognitive recovery than male infants from intracranial hemorrhage associated with prematurity.

Summary

This section examined two forms of developmental abnormalities: autism and schizophrenia. In both cases it is possible to link the behavioral manifestations of the disorder to anomalies in neurobiological systems and to identify a genetic component. In both disorders, too, affected individuals do not show localized damage in circumscribed brain regions, but show evidence of cellular disturbance early in brain development (i.e., during neuromigration), which gives rise to pervasive and dysfunctional reorganization of critical neural systems underlying complex behaviors. Emerging evidence suggests that similar mechanisms may characterize other developmental disabilities, including dyslexia, mental retardation, and attention deficit disorders. Current findings also support the existence of sex differences in brain development and organization, which may in turn differentially affect the response of the brain to early injury.

PATHOLOGICAL MANIFESTATIONS OF COGNITIVE AGING: DEMENTIA

Dementia is a generic term that refers to debilitating deterioration in more than one domain of cognitive function. The involvement of multiple capacities distinguishes dementia from other disorders, such as amnesia and aphasia, that affect a single functional domain (memory and language, respectively). Approximately 50 disorders are known to cause dementia. Most of these are progressive in nature, increasing in severity over time. The age at onset and the rate of progression of symptoms differ dramatically among the major dementing disorders. Most have an insidious onset and develop slowly, sometimes over a period of many years. These include Alzheimer's disease, Huntington's disease, frontotemporal dementia, and the dementia observed in a subpopulation of patients with

Parkinson's disease. The rare neurodegenerative disorder Creutzfeldt–Jakob disease also develops insidiously, but is distinguished by a very rapid rate of progression, often spanning a year or less from the onset of dementia to death. Vascular dementia (or multi-infarct dementia) follows still another pattern of decline. Initial cognitive symptoms develop acutely, but in this case the clinical course typically proceeds in a step-wise fashion over many years, with periods of relative stability punctuated by abrupt deterioration. Age is a major risk factor for dementing illnesses, and the following sections focus on the most prevalent example; Alzheimer's disease.

Alzheimer's Disease Is the Most Common Cause of Dementia

In 1906, the German scientist Alois Alzheimer reported a case study of a woman in her late 40s who developed memory impairment and deficits in other mental faculties that progressively worsened in severity until her death in her early 50s. Subsequent neuropathological examination of this patient's brain revealed grossly apparent and widespread neocortical atrophy that Alzheimer concluded was inconsistent with any previously described disease. Although the disorder that would come to bear his name was originally thought to be quite rare, Alzheimer's disease is now recognized as the most common cause of dementia, accounting for an estimated 50% of all cases, and the fifth leading cause of death among those over 65 years of age. Current estimates are that more than 5 million people in the United States, and 26 million people worldwide, suffer from Alzheimer's disease. The prevalence of the disease is tightly coupled to age and increases after the fourth decade of life. The disease affects about 12% of people over the age of 65, and by 85, as many as 40% are afflicted. Demographic trends indicate that the elderly comprise the fastest growing segment of the population, and as a consequence, the number of people suffering the devastating impact of Alzheimer's disease is projected to nearly triple by 2050. Not surprisingly, efforts to understand the basis of this and related disorders are among the most active research areas in neuroscience.

Memory Impairment Is a Central Feature of Alzheimer's Disease

Alzheimer's original case study provided an accurate description of the clinical course for many patients with this disease. Impairment in establishing and maintaining memories for recent events is frequently the initial sign that brings the patient to the attention of health care professionals. At this early stage, memory for the remote past is relatively preserved, as are other forms of memory that do not require medial temporal lobe structures. Thus, the neuropsychological profile of the Alzheimer's patient provides important clues about the neuropathological progression of this disorder, suggesting that the systems responsible for declarative memory are especially vulnerable. Neuroanatomical studies of the diagnostic hallmarks of Alzheimer's disease (described in the next section) confirm this proposal. Progression of the disorder is marked by inexorable decline across multiple cognitive capacities. Although generally alert and responsive in early and middle phases of the disorder, Alzheimer's patients often experience difficulty identifying the meaning of simple words, uses for common household objects, or the meaning of numbers. Other disturbances can include general confusion, agitation, delusions, social disinhibition, and paranoia. Language abilities can deteriorate progressively, and at the end stage of disease, patients are often mute and densely amnesic for both recent and more remote events. As might be predicted on the basis of this broad profile of impairment, the distribution of neuropathology in individuals dying with late-stage Alzheimer's disease is extensive, involving widespread areas of the limbic system and association cortices, along with many subcortical brain regions, including the basal forebrain cholinergic system, striatum, thalamus, and cerebellum.

Amyloid-Containing Plaques and Neurofibrillary Tangles Are Diagnostic Hallmarks of Alzheimer's Disease That Preferentially Target Memory-Related Brain Regions

There are two types of neuropathological lesions that, when present in sufficiently high numbers, are the diagnostic hallmarks of Alzheimer's disease (Fig. 43.6). Plaques comprise extracellular deposits of amyloid β-protein (Aβ) that, in their mature form, are typically surrounded by a

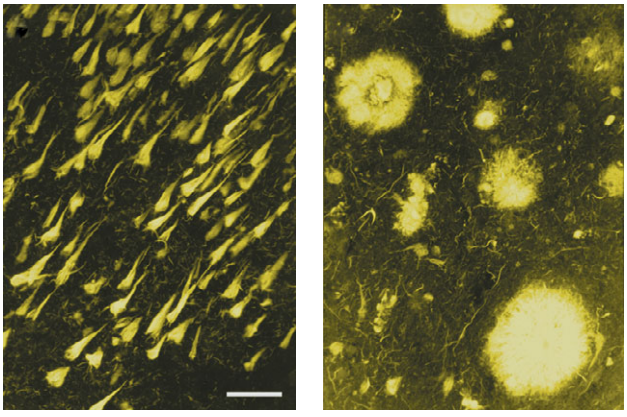


FIGURE 43.6 Photomicrographs of neurofibrillary tangles (A) and amyloid-containing plaques (B) in the hippocampal formation of a patient that died with late-stage Alzheimer's disease. Note that the size and morphological characteristics of plaques in B vary widely. Scale bar: 50 μm. *Courtesy of P. Hof, Mount Sinai School of Medicine.*

shell of dystrophic neurites (see Fig. 43.6B). The neuritic component of plaques consists of morphologically abnormal dendrites and axons of multiple neuronal types, typically associated with substantial numbers of glial cells. The amyloid core of plaques can contain multiple species of Aβ, but prominently includes a form ending at amino acid 42 that is prone to aggregation (Aβ_{42}), and a slightly shorter species (Aβ_{40}) that is normally produced more abundantly in neurons. Plaques vary substantially in size, and although the time it takes for mature neuritic plaques to form is not known precisely, they are likely to evolve slowly, over many years.

In contrast to extracellular plaques, the second neuropathological hallmark of Alzheimer's disease is localized intracellularly. Termed neurofibrillary tangles, these are composed of densely packed, abnormal fibers that occupy much of the cytoplasm and proximal neuronal processes of affected cells (Fig. 43.6A). Indeed the density of these fibers in many tangle-bearing neurons is sufficiently high that the cell body appears swollen and the nucleus displaced. The term "paired helical filaments" (PHF), sometimes used to describe this type of pathology, derives from electron microscopic evidence that tangles consist of pairs of twisted filaments with a highly regular periodicity. Studies using biochemical and immunocytochemical methods have demonstrated that PHF are composed of a cytoskeletal protein found in normal, healthy neurons: the microtuble-associated protein tau. Cytoskeletal abnormalities can disrupt intracellular trafficking of materials that are critical for cell viability, culminating in neuron death. Defining the mechanisms that initiate the hyperphosphorylation of tau and its aggregation into mature neurofibrillary pathology is a prominent avenue of research on Alzheimer's disease.

Investigations aimed at mapping the neuroanatomical distribution of pathology during early and later stages of Alzheimer's disease help account for the progression of clinical symptoms in patients with this disorder. Early in the course of the disease, when memory impairment is often the predominant complaint, plaques and tangles are particularly abundant in the entorhinal cortex and associated projection targets in the hippocampus. Indeed, in many cases virtually all of the neurons that originate the perforant path projection to the hippocampus exhibit neurofibrillary tangles. Emerging evidence also suggests the balance of excitatory and inhibitory drive is disrupted in vulnerable circuitry early in the course of the disease, perhaps contributing to Aβ deposition. In this way the pathophysiology of Alzheimer's disease appears to initially target cortical networks critical for normal memory. As the severity of cognitive impairment progresses, the distribution of neuropathology becomes more widespread, affecting cortical and subcortical systems that support language, semantic knowledge, abstract reasoning, and other capacities.

Genetic Contributions to Alzheimer's Disease

Important clues about the basis of Alzheimer's disease have come from the study of people afflicted with an "early onset" form of the disorder that emerges before the age of 65. This relatively rare and aggressive condition exhibits an increased prevalence in certain families and is therefore thought to have a significant genetic component. To date, three gene mutations linked to early onset Alzheimer's have been identified: one involving the amyloid precursor protein gene on chromosome 21, another in the presenilin 1 gene on chromosome 14, and the third in the presenilin 2 gene on chromosome 1. Each of these mutations is expressed in an autosomal-dominant fashion such that half the offspring of an affected parent develop the disease. Importantly, the common thread among the gene mutations that cause early onset Alzheimer's disease is that each of them influences processing of the amyloid precursor protein (reviewed in Selkoe, 2001).

Several additional lines of evidence support the suggestion that the deposition of amyloid is central to the pathogensis of Alzheimer's disease. By far the most common form of the disorder (accounting for an estimated 98% of cases) has a relatively late onset, after age 65. Among a number of recently identified candidate genes, one risk factor for late-onset Alzheimer's disease is the allelic composition of the apolipoprotein E gene on chromosome 19 (ApoE)—that is, a gene that codes for a glycoprotein involved in cholesterol transport and metabolism. The ApoE gene has three alleles (designated ApoE2, ApoE3, and ApoE4), and patients with Alzheimer's disease carry the ApoE4 allele with a frequency greatly exceeding individuals without the disease. Moreover, the risk of disease is coupled to the number of copies of this specific allele: individuals with no copies of ApoE4 are less likely than the general population to develop Alzheimer's disease, the presence of one ApoE4 allele increases the disease risk four times, and people with two copies of ApoE4 are eight times more likely than the population at large to develop Alzheimer's. Although the precise mechanism by which ApoE4 mediates disease susceptibility is unknown, one hypothesis is that it involves the disruption of processes responsible for scavenging and clearing amyloid peptide from the extracelluar space in brain.

Additional findings relevant to the role of amyloid in Alzheimer's disease come from research on Down's syndrome. Down's syndrome results from the presence of an extra copy of the same chromosome that carries the gene for the amyloid precursor protein; chromosome 21. Thus, it is of considerable interest that people with Down's syndrome who survive until the fourth decade of life

invariably develop a distribution of amyloid plaques in the brain similar to Alzheimer's patients. To the degree that aberrant processing or clearance of amyloid comprises the common pathogenic mechanism of Alzheimer's disease, current findings encourage the view that interventions targeting this pathway might realize an effective treatment for the most common cause of age-related dementia.

Prospects for the Treatment of Alzheimer's Disease

Although Alzheimer's disease is presently incurable, several pharmacological treatments have been approved for the clinical management of cognitive impairment associated with the disorder. All of these treatments target either acetylcholine or glutamate signaling, and while they offer modest, transient benefit for some Alzheimer's patients, they fall far short of an effective cure or prevention. Preclinical progress toward the latter goal includes evidence that immunization with Aβ can both attenuate spatial learning impairment and blunt the progression of amyloid deposition in transgenic mouse models of Alzheimer's disease (reviewed in Selkoe, 2001). Translating this approach to clinical practice faces substantial challenges, and initial attempts indicate that parallel treatment in people can induce life threatening cerebral inflammation. Nonetheless, the recent development of techniques for imaging brain amyloid deposition in vivo has generated considerable excitement, suggesting that it may soon be possible to identify at risk individuals very early in the course of disease, prior to the onset of cognitive decline, when intervention is most likely to be beneficial. The hopeful outlook based on these findings is that vaccine therapies might ultimately yield an effective route for the treatment and prevention of Alzheimer's disease.

Summary

This chapter has considered the neurobiology of cognitive development and aging. Early cognitive development is marked by the emergence of a remarkably complex set of adaptive capacities, including language skills, reasoning abilities, attention, declarative memory, and executive functions. The vast majority of neurons—the essential building blocks of the central nervous system—are present at birth in humans, and rather than depending on the recruitment of new neurons, cognitive development appears to involve the fine tuning and establishment of appropriate circuitry among existing cells. Given the complexity of orchestration required, it is not surprising that many major disorders affecting cognitive function appear to result from a disruption of normal neurodevelopmental events. Higher order cognitive capacities are also vulnerable to deterioration during normal aging. Whereas widely distributed neuron death was previously presumed to underlie these deficits, it now appears that regionally selective changes in connectivity and cell biological function in critical circuits are the causative factors. Aging is also associated with some of the most devastating pathological conditions affecting cognitive function, including Alzheimer's disease. The dramatic progress realized in recent years has, for the first time, established a basis for the rational development of strategies for disease prevention and treatment.

References

Alvarado, M. C., & Bachevalier, J. (2000). Revisiting the maturation of medial temporal lobe memory functions in primates. *Learning & Memory, 7,* 244–256.

Bachevalier, J., & Vargha-Khadem, F. (2005). The primate hippocampus: Ontogeny, early insult and memory. *Current Opinion in Neurobiology, 15,* 168–174.

Baron-Cohen, S., Ring, H. A., Wheelwright, S., Bullmore, E. T., Brammer, M. J., Simmons, A., & Williams, S. C. R. (1999). Social intelligence in the normal and autistic brain: An fMRI study. *The European Journal of Neuroscience, 11,* 1891–1898.

Bauer, P. J. (2002). Long-term recall memory: Behavioral and neurodevelopmental changes in the first 2 years of life. *Current Directions in Psychological Scienc, 11,* 137–141.

Benes, F. M., Turtle, M., Khan, Y., & Farol, P. (1994). Myelination of a key relay zone in the hippocampal formation occurs in the human brain during childhood, adolescence, and adulthood. *Archives of General Psychiatry, 51,* 477–484.

Bizon, J. L., & Gallagher, M. (2005). More is less: Neurogenesis and age-related cognitive decline in Long-Evans rats. *Science of Aging Knowledge Environment, 7,* re2.

Brody, H. (1970). Structural changes in the aging nervous system. *Interdisciplinary Topics in Gerontology, 7,* 9–21.

Burke, S. N., & Barnes, C. A. (2006). Neural plasticity in the ageing brain. *Nature Reviews Neuroscience, 7,* 30–40.

Chugani, H. T., Phelps, M. E., & Mazziotta, J. C. (1993). Positron emission tomography study of human brain functional development. In M. H. Johnson (Ed.), *Brain development and cognition: A reader* (Chap. 8, pp. 125–143). Cambridge, MA: Blackwell.

Conel, J. L. (1939–1963). *The postnatal development of the human cerebral cortex* (Vols. I–VI). Cambridge, MA: Harvard University Press.

de Haan, M., Mishkin, M., Baldeweg, T., & Vargha-Khadem, F. (2006). Human memory development and its dysfunction after early hippocampal injury. *TINS, 29,* 374–381.

Diamond, A. (1990). The development and neural bases of memory functions as indexed by the AB and delayed response tasks in human infants and infant monkeys. *Annals of the New York Academy of Sciences, 608,* 637–676.

Fletcher, B. R., & Rapp, P. R. (2012). Normal neurocognitive aging. In *Handbook of psychology: Biological psychology and neuroscience* (2nd ed., Vol. 3). In press.

Geinisman, Y., de Toledo-Morrell, L., Morrell, F., Persina, I. S., & Rossi, M. (1992). Age-related loss of axospinous synapses formed by two afferent systems in the rat dentate gyrus as revealed by the unbiased stereological disector technique. *Hippocampus, 2,* 347–444.

Giedd, J. N., Vaituzis, A. C., Hamburger, S. D., Lange, N., Rajapakse, J. C., Kaysen, D., … Rapoport, J. L. (1996). Quantitative MRI of the temporal lobe, amygdala, and hippocampus in normal human development: Ages 4–18 years. *Journal of Comparative Neurology, 366,* 223–230.

Hedden, T., & Gabrieli, J. D. (2005). Healthy and pathological processes in adult development: New evidence from neuroimaging of the aging brain. *Current Opinion in Neurology, 18,* 740–747.

Huttenlocher, P. R. (1990). Morphometric study of human cerebral cortex development. *Neuropsychologia, 28,* 517–527.

Huttenlocher, P. R., & De Courten, C. (1987). The development of synapses in striate cortex of man. *Human Neurobiology, 6,* 1–9.

Klingberg, T. (2008). White matter maturation and cognitive development during childhood. In C. A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (2nd ed., Chap. 14, pp. 237–243). Cambridge, MA: MIT Press.

Lenroot, R. K., & Giedd, J. N. (2006). Brain development in children and adolescents: Insights from anatomical magnetic resonance imaging. *Neuroscience and Biobehavioral Reviews, 30,* 718–729.

Lenroot, R. K., Gogtay, N., Greenstein, D. K., Wells, E. M., Wallace, G. L., Clasen, L. S., … Giedd, J. N. (2007). Sexual dimorphism of brain development trajectories during childhood and adolescence, *Neuroimage, 36,* 1065–1073.

Lenroot, R. K., Schmitt, J. E., Ordaz, S. J., Wallace, G. L., Neale, M. C., Lerch, J. P., … Giedd, J. N. (2009). Differences in genetic and environmental influences on the human cerebral cortex associated with development during childhood and adolescence. *Human Brain Mapping, 30,* 163–174.

Morrison, J. H., & Hof, P. R. (1997). Life and death of neurons in the aging brain. *Science, 278,* 412–419.

Ofen, N., Kao, Y. -C., Sokol-Hessner, P., Kim, H., Whitfield-Gabrieli, S., & Gabrieli, J. D. E. (2007). Development of the declarative memory system in the human brain. *Nature Neuroscience, 10,* 1198–1205.

Peters, A., Morrison, J. H., Rosene, D. L., & Hyman, B. T. (1998). Are neurons lost from the primate cerebral cortex during normal aging? *Cerebral Cortex, 8,* 295–300.

Pelphrey, K., & Carter, E. J. (2008). Brain mecahnisms for social perception: Lessons from autism and typical development. *Annals of the New York Academy of Sciences, 1145*, 283–299.

Rapp, P. R., & Gallagher, M. (1996). Preserved neuron number in the hippocampus of aged rats with spatial learning deficits. *Proceedings of the National Academy of Sciences of the United States of America, 93,* 9926–9930.

Rapp, P. R., Deroche, P. S., Mao, Y., & Burwell, R. D. (2002). Neuron number in the parahippocampal region is preserved in aged rats with spatial learning deficits. *Cerebral Cortex, 12,* 1171–1179.

Raznahan, A., Toro, R., Daly, E., Robertson, D., Murphy, C., Deeley, Q., … Murphy, D. G. (2010). Cortical anatomy in autism spectrum disorder: an in vivo MRI study on the effect of age. *Cerebral Cortex, 20,* 1332–1340.

Raznahan, A., Lee, Y., Stidd, R., Long, R., Greenstein, D., Clasen, L., … Giedd, J. N. (2010). Longitudinally mapping the influence of sex and androgen signaling on the dynamics of human cortical maturation in adolescence. *Proceedings of the National Academy of Sciences, 107,* 16988–16993.

Rosenzweig, E. S., & Barnes, C. A. (2003). Impact of aging on hippocampal function: Plasticity, network dynamics, and cognition. *Progress in Neurobiology, 69,* 143–179.

Schore, A. N. (1996). The experience-dependent maturation of a regulatory system in the orbital prefrontal cortex and the origin of developmental psychopathology. *Development and Psychopathology, 8,* 59–87.

Selkoe, D. J. (2001). Alzheimer's disease: Genes, proteins, and therapy. *Physiological Reviews, 81,* 741–766.

Seress, L. (2008). Pre- and postnatal morphological development of the human hippocampal formation. In C. A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (2nd ed., Chap. 12, pp. 187–211). Cambridge, MA: MIT Press.

Shaw, P., Kabani, N. J., Lerch, J. P., Eckstrand, K., Lenroot, R., Gogtay, N., … Wise, S. P. (2008). Neurodevelopmental trajectories of the human cerebrtal cortex. *The Journal of Neuroscience, 28,* 3586–3594.

State, M. W. (2010). The genetics of child psychiatric disorders: Focus on autism and tourette syndrome. *Neuron, 68,* 254–269.

Wang, M., Gamo, N. J., Yang, Y., Jin, L. E., Wang, X. J., Laubach, M., … Arnsten, A. F. (2011). Neuronal basis of age-related working memory decline. *Nature, 476*, 210–213. doi: 10.1038/nature10243. [Epub ahead of print]

Weinberger, D. R., Berman, K. F., Suddath, R., & Torrey, E. F. (1992). Evidence of dysfunction of a prefrontal-limbic network in schizophrenia: A magnetic resonance imaging and regional cerebral blood flow study of discordant monozygotic twins. *The American Journal of Psychiatry, 149,* 890–897.

Wilson, I. A., Gallagher, M., Eichenbaum, H., & Tanila, H. (2006). Neurocognitive aging: Prior memories hinder new hippocampal encoding. *TINS, 29,* 662–670.

Zeamer, A., & Bachevalier, J. (2010). Developmental trajectory of object recognition memory in infant rhesus monkeys with and without neonatal hippocampal lesions. *The Journal of. Neuroscience, 30,* 9157–9165.

Suggested Readings

Bearden, C. E., Meyer, S. E., Loewy, R. L., Niendam, T. A., & Cannon, T. D. (2006). The neurodevelopmental model of schizophrenia: Updated. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology* (2nd ed., vol. 3, pp. 542–569). New York, NY: John Wiley & Sons.

Burke, S. N., & Barnes, C. A. (2010). Senescent synapses and hippocampal circuit dynamics. *TINS, 33,* 153–161.

Casey, B. J., Tottenham, N., Liston, C., & Durston, S. (2005). Imaging the developing brain: What have we learned about cognitive development? *TRENDS in Cognitive Sciences, 9,* 104–110.

Dawson, G., & Toth, K. (2006). Autism spectrum disorders. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology* (2nd. ed., Vol. 3, pp. 318–357). New York, NY: John Wiley & Sons.

Dean, R. L., & Bartus, R. T. (1988). Behavioral models of aging in nonhuman primates. In L. L. Iversen, S. D. Iversen, & S. H. Snyder (Eds.), *Handbook of psychopharmacology* (Vol. 20, pp. 352–392). New York, NY: Plenum.

Deboysson-Bardies, B., de Schonen, S., Jusczyk, P., McNeilage, P., & Morton, J. (1993). *Developmental neurocognition: Speech and face processing in the first year of life*. Boston: Kluwer Academic.

Gallagher, M., & Rapp, P. R. (1997). The use of animal models to study the effects of aging on cognition. *Annual Review Psychology, 48,* 339–370.

Hernandez, D. G., Nalls, M. A., Gibbs, J. R., Arepalli, S., van der Brug, M., Chong, S., … Singleton, A. B. (2011). Distinct DNA methylation changes highly correlated with chronological age in the human brain. *Human Molecular Genetics, 20,* 1164–1172.

Keshavan, M. S., & Murray, R. M. (1997). *Neurodevelopment and adult psychopathology*. Cambridge: Cambridge Univ. Press.

Lenroot, R. K., & Giedd, J. N. (2011). Annual research review: Developmental considerations of gene by environment interactions. *Journal of Child Psychology and Psychiatry, 52,* 429–441.

Mastroeni, D., Grover, A., Delvaux, E., Whiteside, C., Coleman, P. D., & Rogers, J. (2011). Epigenetic mechanisms in Alzheimer's disease. *Neurobiology of Aging, 32,* 1161–1180.

Nelson, C. A., & Luciana, M. (2008). *Handbook of developmental cognitive neuroscience* (2nd ed.). Cambridge, MA: MIT Press.

Nickl-Jockschat, T., Habel, U., Maria Michel, M. T., Manning, J., Laird, A. R., Fox, P. T., … Eickhoff, S. B. (2011, June 20). Brain structure anomalies in autism spectrum disorder—a meta-analysis of VBM studies using anatomic likelihood estimation. *Human Brain Mapping*, doi: 10.1002/hbm.21299. [Epub ahead of print].

Paus, T. (2005). Mapping brain maturation and cognitive development during adolescence. *TRENDS in Cognitive Sciences, 9,* 60–68.

Schmitz, C., & Rezaie, P. (2008). The neuropathology of autism: Where do we stand? *Neuropathology and Applied Neurobiology, 34,* 4–11.

Selkoe, D. J. (2011). Resolving controversies on the path to Alzheimer's therapeutics. *Nature Medicine, 17,* 1060–1065.

Sherwood, C. C., Gordon, A. D., Allen, J. S., Phillips, K. A., Erwin, J. M., Hof, P. R., & Hopkins, W. D. (2011). Aging of the cerebral cortex differs between humans and chimpanzees. *Proceedings of the National Academy of Sciences of the United States of America, 108,* 13029–13034.

Stern, Y. (2009). Cognitive reserve. *Neuropsychologia, 47,* 2015–2028.

Stromswold, K. (1995). The cognitive and neural bases of language acquisition. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 855–870.). Cambridge, MA: MIT Press.

Yassa, M. A., Mattfeld, A. T., Stark, S. M., & Stark, C. E. L. (2011). Age-related memory deficits linked to circuit-specific disruptions in the hippocampus. *Proceedings of the National Academy of Sciences of the United States of America, 108,* 8873–8878.

Peter R. Rapp and Jocelyne Bachevalier

This page intentionally left blank

CHAPTER

44

Visual Perception of Objects

THE PROBLEM OF OBJECT RECOGNITION

Accurate perception and recognition of objects is a crucial ability for visually dependent organisms such as humans—but it is also a complex one. As object and viewer move with respect to each other, the images cast by the object on the retinas of the observer can change radically. Not only will image position and size change, but surfaces may be foreshortened, occluded, or newly revealed. Similarly, changes in ambient illumination can reveal new surfaces, hide others in shadow, or introduce spurious edges where light ends and shadow begins. Despite this, human and nonhuman primates recognize objects quite readily. This means that the visual system is able to extract a reasonably constant "representation" for the shapes and identities of meaningful objects despite such transformations. Image processing by the retina, visual thalamus, and primary visual cortex begins this process by extracting features and representing them over progressively larger receptive fields. However, in order to understand how the brain represents the identity of an object—its defining, invariant features of shape, relations among parts, and its significance—one must examine how circuits beyond the primary visual cortex encode and retain crucial information of a more global or "high-level" nature.

Visual processing can be subdivided into low-level, intermediate-level, and high-level components. The initial, low-level stages are those in which the brain extracts information about the physical properties of the stimulus, such as the edges, brightness, color, and motion of objects present in scenes (see Chapter 26). This processing often is described as bottom-up because it can proceed from low (input) levels up through the visual system without feedback from cognitive processes such as memory. Intermediate-level vision involves combining information about these physical properties in order to detect global properties of an object, such as object shape and the orientation of the object in depth. High-level processes include those that endow the stimulus with meaning, leading to object recognition and classification. For example, recognition and classification depend on object knowledge and context, and can therefore be considered top-down because the neural processing is influenced directly and profoundly by cognitive factors. Other top-down mechanisms, such as selective attention, can operate at even early stages of visual processing (see Chapter 46).

This chapter is concerned primarily with the neural basis of the high-level components of object vision. First, historical evidence is reviewed that helped localize the crucial circuitry to the ventral parts of the temporal cortex. Next, from studies in nonhuman primates, the main pathways are described for object perception and recognition. The complex features of temporal cortex neurons involved in object perception are outlined. Neuroimaging and electrophysiological studies are reviewed to show that very similar processing pathways and mechanisms exist in humans. In addition, human studies are reviewed to consider how the brain may treat special categories of objects and the ways in which experience with objects tunes their neural representations. Finally, visual object perception is considered in the broader context of general object knowledge and memory.

© 2013 Elsevier Inc. All rights reserved.

SUBSTRATES FOR OBJECT PERCEPTION AND RECOGNITION: EARLY EVIDENCE FROM BRAIN DAMAGE

Early Studies of Human Patients Pointed to the Inferior Temporal Cortex as a Site for Object Agnosia

During the late nineteenth century, researchers were occupied with locating the visual centers of the brain. In 1881, Hermann Munk reported that while lesions to occipital cortex produced a complete loss of vision, lesions to temporal cortex resulted in something quite different. He found that temporal lobe lesions in monkeys and dogs would cause the animals to ignore food and water, even if hungry and thirsty. However, they could still navigate rooms and avoid obstacles placed in front of them, indicating that their visual acuity was, at least partially, intact (Brown & Schafer, 1888; Munk, 1881). Munk referred to this syndrome as "psychic blindness," now referred to as visual agnosia, to describe the apparent inability of the animal to understand the meaning of the visual stimuli presented.

Visual agnosia has since been demonstrated in humans who have suffered damage to the occipital and temporal cortices. It refers to a set of disorders affecting object recognition, in which elementary visual capacities such as acuity and visual fields are preserved or grossly intact (for review, see Farah, 2004). Writers such as Potzl in 1928 noted the importance of inferior regions of the temporal lobes, along with adjacent ventral occipital cortex, in most cases of visual agnosia. Subsequent studies upheld these general findings and confirmed the central role of the lingual and fusiform gyri (Fig. 44.1). Patients with visual agnosia typically have bilateral (or, less commonly, right-sided) damage to these regions.

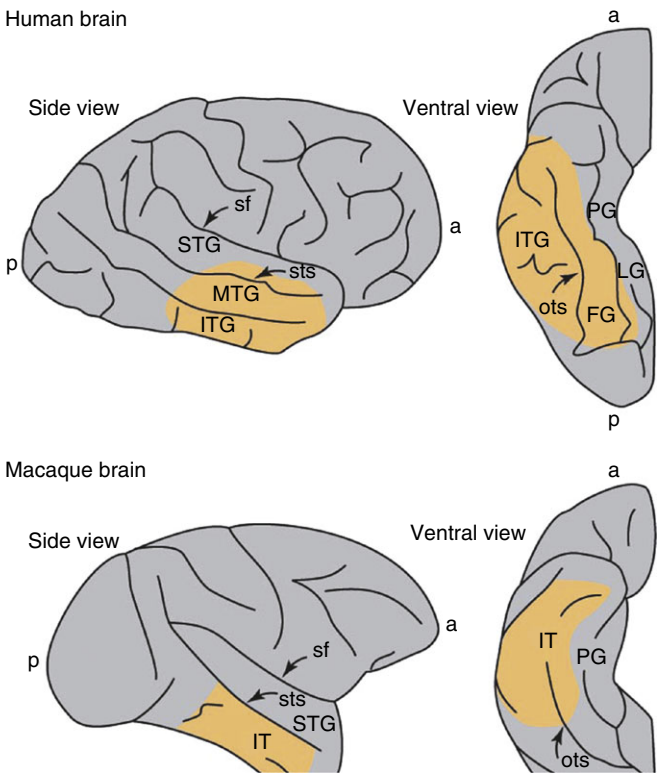


FIGURE 44.1 (Top) Region of the human brain (in shading) within which large bilateral or right-sided lesions cause associative visual agnosia. (Bottom) Region in the macaque monkey (in shading) referred to as the inferior temporal cortex, within which bilateral lesions cause visual discrimination and recognition deficits analogous to agnosic syndromes in humans. a, anterior; p, posterior. FG, fusiform gyrus; ITG, inferior temporal gyrus; LG, lingual gyrus; MTG, middle temporal gyrus; PG, parahippocampal gyrus; STG, superior temporal; ots, occipito-temporal sulcus; sf, Sylvian or lateral fissure; sts, superior temporal sulcus.

Two Types of Visual Agnosia Have Been Distinguished Historically

Two forms of visual agnosia have been characterized. Agnosic patients, whose elementary perception of shape is impaired (despite intact or roughly normal visual acuity) are classified as *apperceptive agnosics* because their impairment is presumed to be at an early stage of the perceptual processing needed for object recognition. Patients suffering from apperceptive agnosia have difficulty with simple discrimination tasks, copying pictures, and cannot name or categorize objects. Patients with such deficits are relatively rare. Usually they have suffered diffuse brain damage, often from carbon monoxide poisoning.

Patients suffering from *associative agnosia* can successfully identify and copy shapes. However, their ability to associate any meaning with the images is severely impaired. Associative agnosia often involves damage to the inferior portions of the posterior cortex, the ventral occipital cortex and/or the lingual and fusiform gyri in the posterior temporal lobe. Associative agnosia is diagnosed according to the following criteria: First, the patient must show difficulty recognizing visually presented objects, as measured by both naming and nonverbal tests of recognition, such as sorting objects by category (e.g., grouping kitchen utensils together, separate from sports equipment) or pantomiming the uses of the object. Second, the patient must show knowledge of the objects through sensory modalities other than vision; that is, be able to recognize items by sound or touch or through verbal questioning ("What is an egg beater?"). Third, the patient must have sensory and perceptual capacity sufficient to allow him or her to describe the appearance of an unrecognized object, to draw it, or identify whether it is the same as or different from a second item. An illustrative case of associative visual agnosia is summarized in Box 44.1. Figure 44.2 shows some copies drawn by this patient, who nevertheless did not recognize the pictures he copied.

BOX 44.1

CASE STUDY OF ASSOCIATIVE AGNOSIA

The subject was a 47-year-old man who had suffered an acute loss of blood pressure with resulting brain damage. His mental status and language abilities were normal, and his visual acuity was 20/30, with a right homonymous hemianopia (blindness in the right visual hemifield). His one severe impairment was an inability to recognize most visual stimuli. For the first three weeks in the hospital, the patient could not identify common objects presented visually and did not know what was on his plate until he tasted it. He identified objects immediately on touching them.

When shown a stethoscope, he described it as "a long cord with a round thing at one end" and asked if it could be a watch. He identified a can opener as a key. Asked to name a cigarette lighter, he said, "I don't know" but named it after the examiner lit it. He said he was "not sure" when shown a toothbrush. He was never able to describe or demonstrate the use of an object if he could not name it. If he misnamed an object, his demonstration of use would correspond to the mistaken identification. Identification improved very slightly when given the category of the object (e.g., something to eat) or when asked to point to named object instead of being required to give the name. When told the correct name of an object, he usually responded with a quick nod and often said, "Yes, I see it now." Then, often he could point out various parts of the previously unrecognized item as readily as a normal subject (e.g., the stem and bowl of a pipe and the laces, sole, and heel of a shoe). However, if asked by the examiner, "Suppose I told you that the last object was not really a pipe, what would you say?" He would reply, "I would take your word for it. Perhaps it's not a pipe." Similar vacillation never occurred with tactilely or aurally identified objects.

After he had spent three weeks on the ward, his object naming ability improved so that he could name many common objects, but this was variable; he might correctly name an object at one time and misname it later. Performance deteriorated severely when the examiner covered any part of the object. He could match identical objects but could not group objects by categories (clothing, food). He could draw the outlines of objects (key, spoon, etc.) that he could not identify.

He was unable to recognize members of his family, the hospital staff, or even his own face in the mirror. Sometimes, he had difficulty distinguishing a line drawing of an animal face from a man's face but always recognized it as a face. The ability to recognize pictures of objects was impaired greatly, and after repeated testing he could name only one or two of 10 line drawings. He was always able to name geometrical forms (circle, square, triangle, cube). Remarkably, he could make excellent copies of line drawings and still fail to name the subject. He easily matched drawings of objects that he could not identity and had no difficulty discriminating between complex nonrepresentational patterns, differing from each other only subtly. He occasionally failed in discriminating because he included imperfections in the paper or in the printers' ink. He could never group drawings of objects by class unless he could first name the subject. Reading, both aloud and for comprehension, was limited greatly. He could read, hesitantly, most printed letters but often misread "K" as "R" and "L" as "T" and vice-versa. He was able to read words slowly by spelling them aloud.

Excerpt from Rubens AB and Benson, D. F. (1971). Associative visual agnosia. Archives of neurology, 24, *305–316.*

Associative Visual Agnosia Involves Deficits in Invariant Object Representations

Closer examination of the manner in which associative agnosics copy and match pictures they cannot identify shows that they are often abnormally attentive to the small details of the image, rather than being guided by the overall shape or "gestalt" of the object itself. This copying is done in a slow, line-by-line fashion, and patients may classify two pictures as different because of a small flaw in the printing. These observations suggest that the associative agnosic patients are unable to match what is in front of them to a unified, stored visual representation of the object. Such stored representations likely include knowledge about the invariant, distinguishing visual attributes (such as the general body shape and relative placement of features of the pig in Fig. 44.2) and/or stored semantic knowledge (the identity of the object as a pig). The underlying problem in visual agnosia could thus result from an inability to generate a high-level representation of the object (a perceptual deficit), a permanent loss of previously stored representations (a memory deficit), or an inability to link the two. In fact, some evidence suggests that different agnosic patients are impaired in distinct stages of visual processing and storage. At the same time, recent work suggests that, at the neural level, "perceptual" and "memory" processes for objects may use the same circuitry (and thus, might be experimentally inseparable), as discussed later in this chapter.

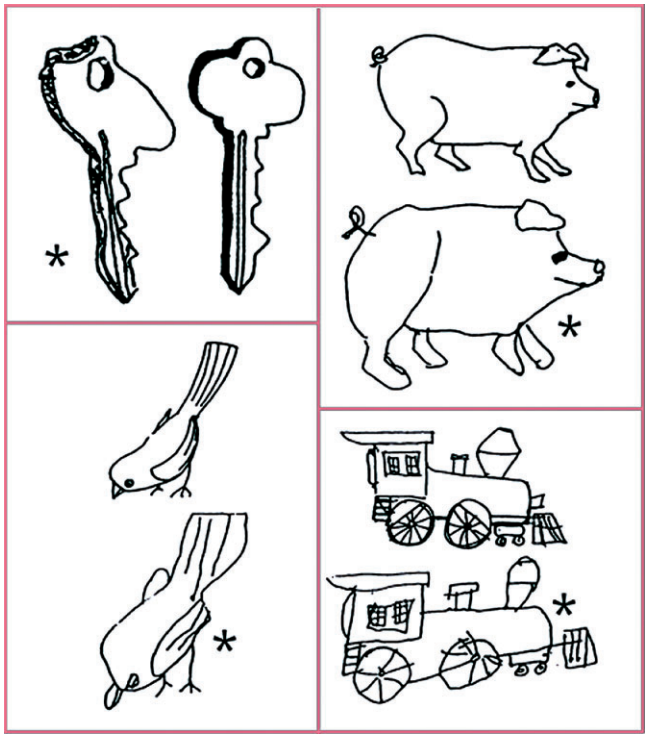


FIGURE 44.2 Pictures that a patient with associative visual agnosia did not recognize but was able to copy. Asterisks indicate the patient's copies. *From Rubens and Benson (1971).*

Object Recognition in Monkeys Depends on the Inferior Temporal Cortex

One of the earliest experimental studies on the role of the temporal cortex in visual recognition was that of Klüver and Bucy, who described in the 1930s what is now known as the Klüver–Bucy syndrome. When these researchers removed the temporal lobes of monkeys bilaterally, the monkeys demonstrated complex changes in visual, social, sexual, and eating behavior. There was significant impairment in their ability to recognize and/or discriminate visual stimuli. These changes in visual ability were not accompanied by any apparent loss of visual acuity. Subsequent research eventually identified the region now known as inferior temporal (IT) cortex as the critical area for producing the visual component of the syndrome. Damage to medially adjacent temporal-lobe regions (i.e., the amygdala) failed to produce visual deficits, but instead led to the social and sexual abnormalities. Monkeys with bilateral IT lesions are impaired both in learning to distinguish between different visual patterns or objects, and in retaining previously acquired visual discriminations. Figure 44.1 shows the location of this area in the macaque brain, relative to the areas implicated in agnosia in humans.

What type of stimulus representation is impaired after IT lesions? A great deal of research has been devoted to this question. Generally, such experiments measure whether monkeys can discriminate different stimuli, before and after experimental lesions in IT cortex—and from any such differences, object-processing mechanisms are inferred. A current summary might be that IT cortex represents aspects of the intrinsic shape of the stimulus that are useful for recognition (e.g., the salient features in a face and their relative positions), generalizing across idiosyncratic aspects of stimulus appearance that depend on specific viewing conditions (see later).

Position in the visual field (retinotopy) is one of these specific visual properties that is independent of object identity, and normal monkeys will easily generalize a visual discrimination learned in one part of the visual field to another. For example, if monkeys learn to tell apart stimuli in one half of visual space, they will easily be able to perform the discrimination when the same stimuli are presented to the other half of visual space. However, monkeys with IT lesions are impaired at this generalization; presumably they have lost representations that are invariant across visual field locations.

Retinal image size of a stimulus is a related visual property that depends on viewing conditions and not just intrinsic object geometry, and this is another perceptual distinction that is difficult for monkeys with IT lesions. For example, monkeys trained to discriminate between two disks of different absolute sizes are unable to perform the discrimination when those disks are presented at variable distances (hence with variable retinal image sizes). Instead, they respond on the basis of retinal size or distance. This result implies that representations within IT cortex normally encode the absolute size of an object, an intrinsic object property useful for recognition, rather than its distance per se or its retinal image size. Finally, variations in illumination prevent monkeys with IT lesions from seeing the equivalence of objects, implying that representations in IT cortex are unaffected by patterns of shadow and light falling on object surfaces.

Summary

Studies of brain damage in humans, and experimental lesions in monkeys, both point to the crucial role of the inferior temporal cortex during object perception and recognition. In both species, analysis of the ensuing deficits suggests that this region participates in creating and storing neural representations of the defining features and invariant aspects that identify a visual object or class of objects.

VISUAL PATHWAYS FOR OBJECT PROCESSING IN NONHUMAN PRIMATES

"What" versus "Where"—Dividing the Visual Cortex of Monkeys into Occipitotemporal and Occipitoparietal Pathways

Early evidence that the visual cortex is divided into separate processing streams was based on the contrasting

effects of inferior temporal and posterior parietal cortex lesions in monkeys. Inferior temporal lesions cause severe deficits in visual discrimination tasks, but they do not affect animals' performance on visuospatial tasks (e.g., visually guided reaching and judging which of two objects lies closer to a visual landmark). In contrast, parietal lesions do not affect visual discrimination ability but instead cause severe deficits on visuospatial performance. On the basis of behavioral results such as these, Ungerleider and Mishkin (1982) proposed the existence of two processing streams: an occipitotemporal or ventral processing stream for mediating the visual recognition of objects ("what" an object is) and an occipitoparietal or dorsal processing stream for mediating the appreciation of the spatial relationships among objects and the visual guidance toward them ("where" an object is). Goodale and Milner (1992) later proposed a modification of this model, emphasizing "perception" versus "action" for ventral and dorsal processing streams, respectively. More recently, it has been proposed that the dorsal stream can be further subdivided into three separate pathways subserving spatial working memory (occipitoparietal-prefrontal), visually guided actions (occipitoparietal-premotor), and navigation (occipitoparietal-medial temporal) (Kravitz, Saleem, Baker, & Mishkin, 2011). The dorsal processing pathways and their importance for spatial cognition are the topics of Chapter 45.

As described in Chapter 26, physiological studies of neuronal properties also support the functional distinction between ventral and dorsal pathways. Neurons within the ventral stream (in distinct modules within V1 and V2, area V4, and inferior temporal areas TEO and TE) respond selectively to visual features relevant for object identification, such as color, shape, and texture. Neurons within the dorsal stream (in other modules within V1 and V2, areas V3, V3A, middle temporal area MT, medial superior temporal area MST, and additional areas in the inferior parietal cortex) instead respond selectively to spatial aspects of stimuli, such as the direction and speed of stimulus motion. Such neurons also respond when the animal visually tracks a moving target. The remainder of this section describes the anatomy of the *occipitotemporal pathway* for object recognition.

The Object Recognition Pathway in Monkeys Consists of an Interconnected Set of Cortical Areas

In both monkeys and humans, cortical regions comprising the object recognition pathway lie directly adjacent to the primary visual cortex (V1) in the occipital lobe, extending progressively into more anterior and ventral portions of the temporal lobe (Fig. 44.3). The cortical analysis of objects begins in V1, where information about contour, orientation, color composition, and brightness is represented in subsets of neurons coding for each point in the visual field (see Chapter 26). Information from V1 projects forward to interdigitating thin, thick, and interstripe regions—within V2 (see Chapter 26). From the thin and interstripe regions in V2 (specialized, respectively, for color and form information, roughly speaking), neural signals proceed forward to area V4 on the lateral and ventromedial surfaces of the hemisphere and to a posterior inferior temporal area just in front of V4, area TEO. From both V4 and TEO, signals related to object form, color, and texture proceed forward to area TE, the last exclusively visual area within the ventral stream for object recognition. Together, areas TEO and TE constitute IT cortex.

Progressing forward along the occipitotemporal pathway from V1 to TE, there is a gradual shift in the nature of connectivity, which is significant for the eventual extraction of invariant object features. At successive stages, patterns of cortical projections become successively less topographic or "point to point." While connections from V1 to V2 maintain topographic order, the organization becomes looser between V2 and V4, and inputs to area TE from V4 and TEO retain no obvious retinotopic organization. Likewise, pathways interconnecting areas of the ventral stream on the two sides of the brain via the corpus callosum and/or the anterior commissure tend to be restricted to the representation of the midline of visual space early in the pathway, but their representation becomes more widespread at higher levels in the pathway. This loss of retinotopy within area TE means that single neurons respond to objects anywhere in the visual field. This evidence suggests that the initial explicit information about spatial position in the visual field ("retinotopy") has been converted to information about object identity, in at least some parts of IT cortex.

However, neural processing in the object recognition pathway is not a simple elaboration of information from lower-order to higher-order areas. At all stages, the neural connections are reciprocal, such that an area receiving feedforward projections from an area earlier in the ventral stream also provides feedback projections to that area. Long-range nonreciprocal connections, such as from area TE back to V1, also exist. On the one hand, feedforward projections provide bottom-up, sensory-driven inputs to subsequent visual areas, and severing these projections disconnects subsequent areas from their visual input. On the other hand, the precise functions of reciprocal, feedback projections are still unknown. Probably these feedback projections play a top-down role in vision (as in selective attention) by modulating activity in earlier areas of the ventral pathway (see Chapter 46). In addition to feedforward and feedback projections, there also appear to be "intermediate-type" projections that link areas at the same level of the visual hierarchy. These are seen most notably between areas of the dorsal and ventral streams. For example, area V4 is interconnected with

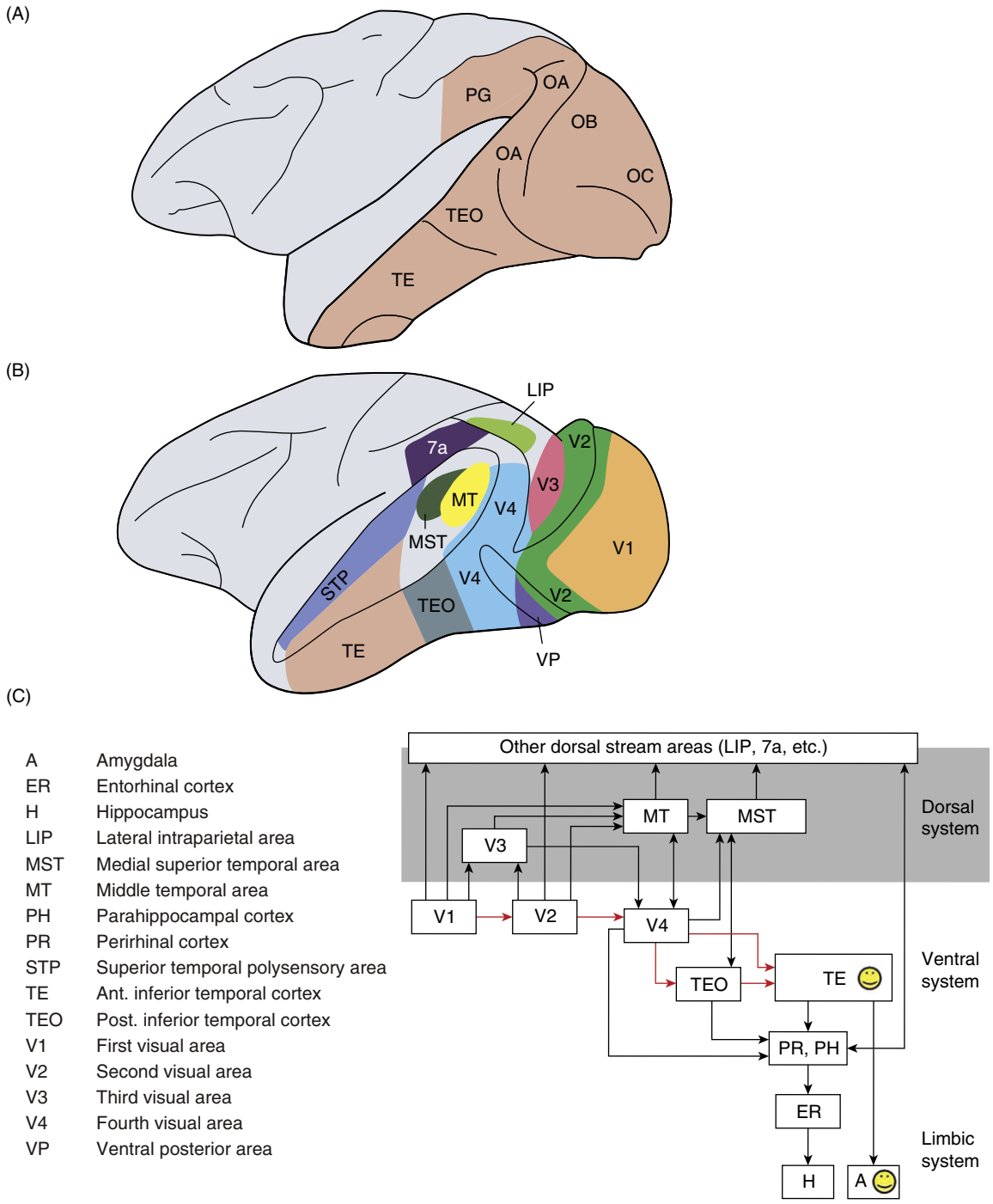


FIGURE 44.3 (A) Schematic diagram of the macaque monkey brain illustrating the two cortical visual system model of Ungerleider and Mishkin (1982). According to this model, there are two major processing pathways, or "streams," in the visual cortex, both originating in the primary visual cortex: a ventral stream, directed into the temporal lobe and crucial for the identification of objects, and a dorsal stream, directed into the parietal lobe and crucial for spatial perception and visuomotor performance. The shaded region indicates the extent of the visual cortex in the monkey, and the labels (OC, OB, OA, TEO, TE, PG) indicate cytoarchitectonic areas according to the nomenclature of von Bonin and Bailey. *Adapted from Mishkin, Ungerleider, and Macko (1983).* (B) A lateral view of the monkey brain illustrating the multiplicity of functional areas within both processing streams. (C) Some of the pertinent connections of the inferior temporal cortex with other cortical areas and medial temporal-lobe structures. Red lines indicate the main afferent pathway to area TE, which includes areas V1, V2, V4, and TEO. For simplicity, only projections from lower-order to higher-order areas are shown, but each of these feedforward projections is reciprocated by a feedback projection. Faces indicate areas in which neurons selectively responsive to faces have been found. *Adapted from Gross, Rodman, Gochin, and Columbo (1993).*

motion-sensitive areas MT and MST. Thus, an anatomical substrate exists for interactions between the two processing streams, and these might serve to integrate what an object is with where it is located.

All areas within the ventral stream also have heavy interconnections with subcortical structures, notably the pulvinar, claustrum, and basal ganglia. In addition, each ventral stream area also receives subcortical modulatory inputs from ascending cholinergic projections from the basal forebrain, and ascending noradrenergic projections from the locus coeruleus. These projections are thought to play a role in the storage of information in the cortex, and the influence of arousal on information processing, respectively. Finally, visual information is sent *from* IT cortex to the most ventral and anterior reaches of the temporal lobe, notably the perirhinal cortex and parahippocampal areas TF and TH. These regions in turn project, via the entorhinal cortex, to medial temporal-lobe structures, such as the hippocampus, which contribute to forming long-term memories of visual objects and their contexts (see Chapter 48). Information is also sent from IT cortex to the prefrontal cortex, which plays an important role in working memory, holding an object briefly in mind when it is no longer visible (see Chapter 50). Finally, there are direct projections from IT cortex to the amygdala, which is important for attaching emotional valence to a stimulus.

Summary

Areas along the occipitotemporal pathway of the monkey are organized hierarchically, proceeding from V1 > V2 > V4 > TEO > TE. At progressively higher stages of this hierarchy, there is a corresponding loss (or conversion) of retinotopic information, such that information about the spatial location of the object is gradually changed. In addition to this bottom-up information (supplied by feedforward projections from lower-order to higher-order visual areas within this pathway), there are also reciprocal, feedback projections from higher-order to lower-order visual areas. These latter projections are thought to play a top-down role in vision.

NEURONAL PROPERTIES WITHIN THE OBJECT RECOGNITION PATHWAY

Neurons along the Occipitotemporal Pathway are Sensitive to Increasingly More Complex Physical Features of Objects

The different cortical areas in the occipitotemporal pathway share a number of physiological characteristics. Consistent with a role in object recognition, all areas in the pathway contain populations of neurons sensitive to the shape, color, and/or texture of visual stimuli. However, at progressively higher levels, neurons have larger receptive fields and their stimulus selectivity is more complex to characterize. Thus, higher-order properties (consistent with invariant representation of objects) usually are attributed to neurons in higher tier areas. For instance, many V1 neurons function as local spatial filters, signaling the presence of contours at particular positions and orientations in the visual field. In contrast, an increasingly higher proportion of neurons in higher tier areas (e.g., V2 and beyond) apparently respond to *illusory* contours (i.e., contours implied by stimulus context and higher-order properties, not due to simple light-dark contrast) across increasingly larger regions of the visual field.

Additional higher-order, object-related properties are also seen at higher tier cortical levels. One of these properties is increasing stimulus invariance (see below)—namely, similar shape selectivity despite variations in the type of lower-order cues used to generate the shape (such as luminance, color, motion, etc.). Another of these properties is increasingly strong, silent suppressive zones surrounding the classical, excitatory receptive field of neurons. These zones are called silent because visual stimulation of them normally does not cause any change from the baseline activity of the neurons—although it can nevertheless powerfully affect responses to a stimulus placed in the classical receptive field. For example, many V4 neurons respond maximally to a stimulus only if it stands out from its background based on differences in form—as in the well-known perceptual phenomenon of figure/ground segregation, a computation that remains difficult to achieve in machine-based object recognition.

Farther forward along the occipitotemporal pathway there are further increases in the complexity of the critical features needed to activate many neurons. Connor and colleagues have demonstrated that neurons in V4 and the posterior portion of IT cortex (area TEO and posterior area TE) are responsive to single contours as well as particular combinations of contours (Pasupathy & Connor, 2002; Brincat & Connor, 2006). These neurons are often insensitive to the position of the contour(s) within a more complex object, indicating that they can participate in the coding of many different complex objects. The convergence of outputs from such neurons could potentially result in neurons selective for objects of even further complexity—which is precisely what is observed within anterior portions of the IT cortex. Specifically, neurons in the anterior two-thirds of IT cortex (area TE) are maximally activated by more complex arrangements of features, including intact objects and even faces.

At these highest levels, it can become very difficult to discover the optimal stimulus for a given single cell by simply testing various visual stimuli because the number of possible test stimuli is infinite. This experimental problem is related to the idea, developed sometime in the late

1960s, that every object would be coded by maximal firing of a single neuron in IT cortex; this was termed the "grandmother neuron" hypothesis because each "grandmother" was thought to cause maximal firing in a single neuron (Gross, 2002). Theoretically, if this neuron were to be destroyed, we should lose the ability to recognize our grandmother. This idea, which was controversial from its inception, is considered by most researchers to be overly simplistic: there simply are not enough neurons to represent all the different objects we can recognize.

Instead, a "population code" is more likely, in which each specific object (and perhaps each viewpoint; see Box 44.2) causes a correspondingly unique pattern of firing in multiple neurons that share connections and overlapping functional selectivity. This latter idea is supported by evidence suggesting that IT neurons are anatomically grouped into functionally similar patches, columns, or areas. Based on findings from single-unit and optical recording, Tanaka (1996) has suggested that IT neurons that respond to common visual features are grouped together into cortical columns that run perpendicular to the cortical surface (Fig. 44.4). This object-column model is similar to that for other well-known columns in early visual areas, which are selective either for orientation, ocular dominance, spatial frequency, or direction of motion. However in Tanaka's model, there is a continuous mapping of complex feature space within a larger region that contains several partially overlapping columns (Fig. 44.4), rather than just one periodic feature, such as orientation, mapped repeatedly. Newer evidence from functional magnetic resonance imaging in awake monkeys further supports the notion that neurons with common functional properties are grouped together within temporal cortex (Op de Beeck, Deutsch, Vanduffel, Kanwisher, & DiCarlo, 2008).

Thus, a given object is not represented by the activity of a single neuron, but by the activity of many neurons across different columns or regions (i.e., "population coding"). A significant benefit of combination coding over a sparse coding scheme (i.e., "The Grandmother Cell Hypothesis") is that it is much more flexible. It can provide robustness to changes in input and precision of representation. Although the image of an object projected on the retina changes in response to variations in illumination and viewing angle, the pattern of activity in IT cortex will be largely maintained, insofar as the clustering of neurons with overlapping but slightly different selectivity serve as a buffer to absorb such changes.

Invariant Representations in the Occipitotemporal Pathway

As described above, the precise nature of the visual information regarding a particular object reaching the retina can change dramatically due to changes in its position, distance, illumination, and orientation relative to the viewer. However, the identity of the object under examination doesn't change. The function of the occipitotemporal pathway can be defined in large measure as determining and encoding those *invariant* features of objects that are useful for recognition.

As one progresses forward from V1 to TE, receptive fields become larger and the retinotopic maps less precise. By area TE, receptive fields are large and bilateral, typically encompassing the central 20° or more—up to the entire visual field in far anterior TE. The large receptive fields of TE neurons signal the presence of a given feature, collection of features, or object, regardless of where they fall in the visual field. In other words, they exhibit invariance of stimulus coding across changes in retinal position (i.e., "position invariance"). Indeed, many neurons in TE show maintained selectivity for their "preferred" stimuli over more complex kinds of transformations, such as size, distance from the subject in depth, or degree of ambient illumination; thus, they show invariance over a number of the changes that normally take place as a moving object is viewed under naturalistic conditions.

Recent experiments by Li and DiCarlo (2008; 2010) have shown how stimulus invariance might be achieved through an associative learning process based on repeated exposures to the same stimulus under viewing conditions that change in a predictable way. Under normal viewing conditions, movement of an object (caused either by the object moving in space and/or by the eye moving relative to the object) will result in a rapid shift in the location of that object on the retina—and yet the identity of the object has not changed. Through repeated exposures, the brain can theoretically learn to associate these different retinal images as the same object in different positions, thus generating a *position invariant* representation. To test this hypothesis, Li and DiCarlo recorded from neurons in IT cortex while simultaneously presenting monkeys with an object at a fixed position from where the monkeys were currently looking. When the monkeys would move their eyes, the object would change to a new object. They predicted that if such an associative learning process were responsible for generating invariant representations, this manipulation should result in these two stimuli being confused as the same object—which was indeed the case. Initially, the neuron responded to the first object only and not the second. However, over time it responded to both, suggesting that the neuron perceived the two stimuli as the same. The ability of downstream neurons to discriminate between the two stimuli had been abolished. Subsequent experiments demonstrated the same effect for object size, suggesting that a similar mechanism might be responsible for generating *size invariance.*

BOX 44.2

THE ISSUE OF VIEWPOINT DEPENDENCE IN OBJECT RECOGNITION

Although we know something of the underlying anatomy and physiology, the mechanisms by which object recognition is accomplished by the brain have been a matter of heated debate. The pioneering theoretical work of Marr in the late 1970s set the stage for current theories, which can be divided into two main groups: object-based (view independent), and image-based (view dependent) theories.

Object-based theories propose that objects are represented in the brain as hierarchies of parts that are, to a large degree, independent of any particular viewpoint. One of the best known examples is Biederman's (1987) recognition by components (RBC), which postulates that objects can be defined as a set of 30 or so primitive shapes, called *geons* (Fig. 44.B1). In this theory, geons are three-dimensional component building blocks that, when combined, can describe any arbitrary object. For example, the telephone in Figure 44.B1 can be constructed from geons 1, 3, and 5. Critically, these geons were thought to be distinguished from one another on the basis of five properties: curvature, collinearity, symmetry, parallelism, and cotermination—properties integral to gestalt theory (Koffka, 1935). Neurophysiological support for object-based models of object recognition was supplied by researchers such as Tanaka and colleagues, who identified neurons in the temporal cortex of macaques selective for decomposed versions of real-world objects (Tanaka, Saito, Fukada, & Moriya, 1991; Tanaka, 1996).

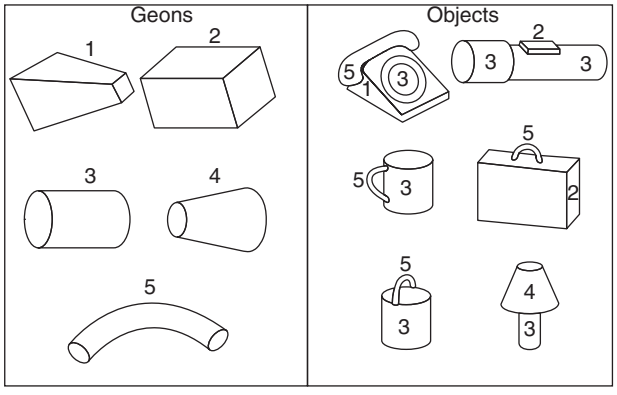


FIGURE 44.B1 Illustration of geons and how they are arranged to form objects. (Left) A given view of an object can be represented by an arrangement of simple primitive volumes, or geons, five of which are shown here. (Right) Only two or three geons are required to uniquely specify an object. The relations among the geons matter, as illustrated with the pail and cup. *From Biederman (1990).*

Object-based theories are considered to be "view-independent" because they rely on hard-coded representations of geons (or similar primitives) encompassing all possible orientations. *Image-based* (also known as *view-dependent*) theories, however, suggest that recognition is linked to the appearance of objects as originally seen. In other words, the brain is not thought of as storing a central, invariant representation of an object, but instead as storing information from multiple independent views (e.g., a view of the telephone as in the figure, but also an end-on or side view, a view from above, etc.). A potential problem with this idea is that it might seem that for reliable object recognition, stored representations for *all* views of objects would need to exist. The solution is to suggest that a small set of characteristic or *prototypical* views would suffice. Proponents of this approach have offered a variety of different theoretical normalization mechanisms for generalizing from unfamiliar views to the familiar, prototypical stored views (Bulthoff, Edelman, & Tarr, 1995). This theory does, however, argue that subjects will be better at recognizing objects when presented in familiar vs. unfamiliar views, which is indeed the case (Bulthoff & Edelman, 1992). This effect has also been demonstrated at the neuronal level. For example, in one study, monkeys were trained to recognize specific novel three-dimensional objects from any viewpoint after experience with a few prototypical views in a discrimination task (Logothetis, Pauls, & Poggio, 1995). After months of such training, the monkeys were found to have neurons in TE that exhibited strong selectivity for precisely those views on which the animal had been trained. These results support the view-dependent account of object recognition.

At first glance, object-based theories appear more compelling, as they attempt at the outset to solve the central problem associated with object recognition, namely reliable recognition despite variation in the precise retinal image. However, one serious challenge is how the brain chooses the correct structural interpretation among the possible ones, as a given set of geons may be compatible with several different objects. Psychophysical testing also indicates that object-centered encoding is not the whole story; studies have shown substantial and robust viewpoint effects even for extremely simple three-dimensional volumes resembling proposed geon primitives.

However, image-based approaches also face challenges. Image-based theories have been most successful in accounting for *exemplar-level* recognition; that is, discriminating

BOX 44.2 *(cont'd)*

between individual visually similar objects within a relatively homogeneous class of objects such as faces, cars, or birds. Though recognition at the exemplar level is often important, there is little question that so called *class-level* recognition actually constitutes much of the job of recognition. For example, for objects such as chairs, telephones, and toasters, we rarely need to discriminate two instances of the class and instead we learn to recognize such objects at a basic category or class level. Image-based models suggest that the brain represents the appearance of a specific individual example of an object from specific viewpoints, and thus seem unsuitable or unnecessary for class-level recognition. Moreover, even for a single exemplar, it is uncertain how many prototypical views are enough for robust recognition. One solution would be to store as many views of an object as possible so that new views would always be similar enough to some previously stored view. However, it is unclear that there can ever be sufficient memory to do so for all the objects that the observer becomes proficient at recognizing under varied conditions.

Face Selective Neurons in IT Cortex Sensitive to Socially Relevant Aspects of Facial Stimuli

One of the more striking characteristics of monkey temporal cortex is the presence of neurons with responses selective for face stimuli. First discovered by Gross, Bender, and Rocha-Miranda (1969), these neurons have received considerable attention, not only for their role in social communication but also for their significance to theories concerning object recognition. Early recording studies reported that face-selective neurons could be found throughout temporal cortex. However, recent studies combining functional neuroimaging with targeted neuronal recordings in nonhuman primates show that face-selective neurons are concentrated into small "patches," located primarily along the lower bank of the superior temporal sulcus and extending onto the adjacent IT gyrus (Tsao, Freiwald, Tootell, & Livingstone, 2006; Bell et al., 2011). It is currently hypothesized that these patches may be homologous to the face-selective regions in human cortex.

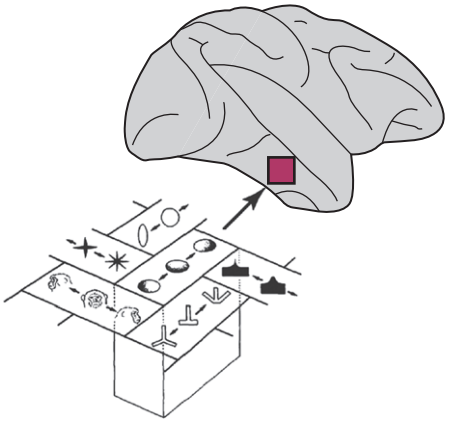


FIGURE 44.4 Schematic diagram illustrating the columnar organization in area TE of the monkey. Neurons with similar but slightly different selectivity tend to cluster in elongated vertical columns perpendicular to the cortical surface. Stimuli shown are examples of the critical features for the activation of single neurons in TE. Most neurons, as shown, require moderately complex features for their activation, although some can be driven by simpler stimuli, such as color, orientation, and texture. *Adapted from Tanaka (1996).*

Face-selective neurons vary in the degree and nature of their preference for face stimuli. Many respond well to both real faces and face pictures, but give little or no response to any other stimuli tested, including other complex objects, texture patterns, and images in which the features making up the face are rearranged or scrambled (Fig. 44.5). Thus, at least some face-selective neurons appear to be sensitive to very general or global aspects of a face, such as the prototypical arrangement of the eyes, nose, and mouth into a face-like configuration.

Other neurons respond to specific face components, such as the presence of eyes *per se*, the distance between the eyes, or the extent of the forehead, and some are sensitive to the direction of gaze of the eyes, which is an important social signal for both monkeys and humans. Other neurons are sensitive to different facial expressions (e.g., open-mouth threat vs. submissive or fearful faces). Finally, features relevant for facial identity discrimination have also been observed at the level of the individual neuron. Although such neurons may merely be sensitive to facial features, they are nonetheless likely to participate in circuits responsible for recognizing a particular individual. Indeed, recent work has demonstrated that electrical microstimulation of face-selective neurons biases object categorization in the direction of face perception, thus establishing a causal relationship between face-selective neurons and face perception (Afraz, Kiani, & Esteky, 2006).

These observations have led to the proposal that the primate temporal lobe has evolved specialized mechanisms for the encoding and recognition of biologically significant stimuli, especially faces. Are such regions innate, or recruited and trained as a result of experience? This is another form of the classical nature-versus-nurture question. Some evidence (e.g., preferential looking at faces in newborns) suggests that face-selectivity may indeed be innate, but experience likely improves such selectivity, as face recognition improves throughout development.

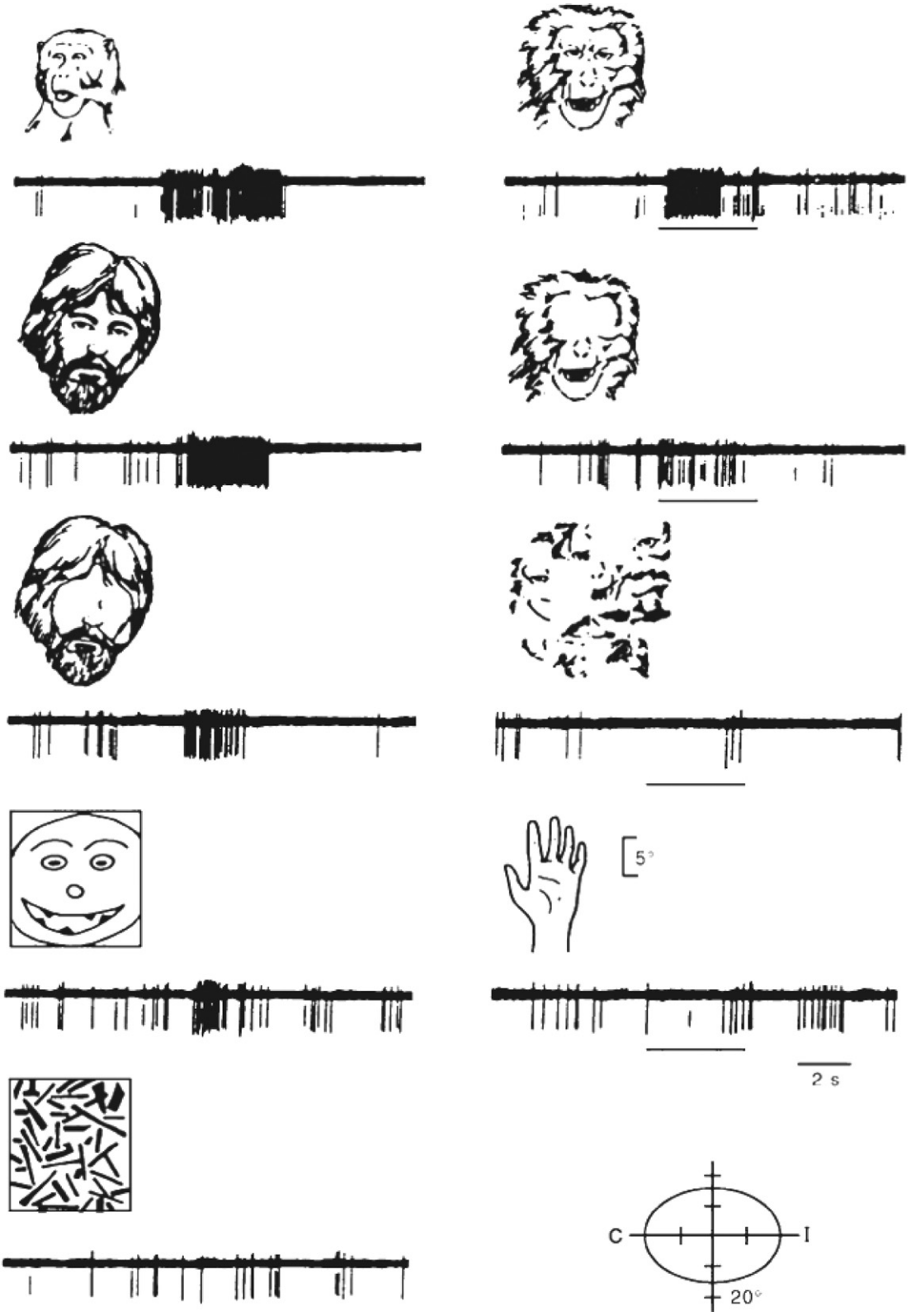


FIGURE 44.5 Activity of a neuron in the superior temporal sulcus that responded better to faces than to all other stimuli tested. Removing eyes on a picture or representing the face as a caricature reduced the response. Cutting the picture into 16 pieces and rearranging the pieces eliminated the response. All the unit records are representative ones chosen from a larger number of trials. The receptive field of the neuron is illustrated on the lower right. C, contralateral; I, ipsilateral visual field. *From Bruce, Desimone, and Gross (1981).*

Response Properties of Neurons in the Object Recognition Pathway are Affected by Experience

The previous sections illustrate how the object recognition pathway, especially area TE, represents features that are useful for recognition. Some neurons respond selectively to the salient features in a face, or their relative positions, while ignoring aspects of stimulus appearance that depend on viewing conditions, such as absolute size and luminance. However, in ecological terms, what is useful and salient within the image of an object or scene changes with experience and context, and these changes are reflected in the way the brain represents stimuli over time. To accommodate this need, neurons in IT cortex and the adjacent perirhinal cortex show effects of both recent viewing history and long-term experience with stimuli.

Experiments by Miller and Desimone (1994) have shown how short-term changes in response properties of neurons in temporal cortex encode aspects of both stimulus recency and short-term stimulus significance. For example, for many of these neurons, responses to an object that is presented repeatedly *decrease* in a systematic fashion, potentially signaling that the object has been seen recently. Such reduced signals may be useful for redirecting attention to novel objects. This kind of response decrement, termed *repetition suppression*, has been found both under anesthesia and for stimuli that an awake animal is trained to ignore. This same effect was later shown to persist over the course of an entire recording session (>1 hour), after hundreds of intervening stimuli. In essence, repetition suppression enables IT neurons to act as filters for novel stimuli. In contrast, if a monkey is trained to do a matching task in which it must hold a particular visual pattern in memory over a short delay, some temporal neurons *increase* their firing rates specifically when the to-be-remembered stimulus is recognized. This phenomenon, termed *mnemonic enhancement*, can be thought of as a mechanism in which bottom-up processing of the stimulus is affected by top-down, cognitive factors, namely, expectation of the relevant stimulus. Stimuli that are repeated but are not behaviorally relevant in the context of the task do not show this enhancement effect.

Experiments such as these demonstrate that the magnitude of neuronal responses can be affected by the experience of the subject with the stimulus and by its significance, at least in the short term. Other studies suggest that the kind of familiarity gained over longer periods of time—akin to the development of "expertise" as discussed later in the chapter—can also be reflected in the responses of TE neurons. For example, in monkeys that have been trained for as long as a year to discriminate between objects, more than 30% of the neurons show selectivity for those objects compared to untrained monkeys in which roughly 10% of the neurons show such selectivity (Kobatake, Wang, & Tanaka, 1998).

Some TE and perirhinal neurons can also represent the *history of association between objects* in the activity patterns they exhibit in response to members of a set of stimuli, as demonstrated by *paired associate* tasks, in which monkeys are trained to associate arbitrary stimuli with one another for reward. In one such experiment, monkeys were first shown one member of a learned pair and then were required to choose the associated "match" from a display of two stimuli while single-unit activity was recorded in anterior IT cortex (Sakai & Miyashita, 1991). Across all neurons recorded, the authors found neurons that responded best to stimuli that had been paired together during training. Neurons in temporal cortex developed conjoint selectivity (*pair coding*) for the members of associated pairs. It was shown that pair coding by TE neurons was decreased by surgically interrupting the feedback projections to them from the perirhinal and entorhinal cortices. Thus, return projections appear to play a specific role in associating objects because the pair-coding property depends on them.

Although associative learning effects typically have been seen in studies after weeks or months of training, there is evidence that small but detectable shifts in neuronal population selectivity occur after only a day or so of experience; this suggests that TE neurons tend to develop pair-coding properties more rapidly than previously thought (Erickson, Jagadeesh, & Desimone, 2000; Messinger, Squire, Zola, & Albright, 2001). Thus, the pattern of selectivity shown by TE neurons at any given moment in time may partly reflect a record of their exposure to stimuli seen together. Overall, TE circuits appear to represent not only complex visual features, but also the history of their meanings and relationships in a behavioral context.

Summary

At successive levels of processing along the occipitotemporal pathway, increasingly complex kinds of information about objects are extracted. In area TE of IT cortex, some neurons are preferentially responsive to faces and other biologically significant stimuli. In addition, TE neurons apparently extract some invariant aspects of objects. For example, TE neurons show similar selectivity for object shape, regardless of large changes in their retinal position, size, distance in depth, and degree of ambient illumination. Finally, as the result of experience, some TE neurons can come to represent novel objects and/or the associative relations between objects. This evidence tells us that neurons in higher-order visual areas are not necessarily limited to responding to a particular, unchanging set of features, but can adapt dynamically to different behavioral objectives, making them ideally suited as the neural substrate for object recognition.

FUNCTIONAL NEUROIMAGING OF OBJECT RECOGNITION

The Human Brain Also Has Ventral and Dorsal Processing Streams

The various visual impairments produced by focal lesions in clinical cases suggest that the human visual cortex, like that of the monkey, contains two anatomically distinct and functionally specialized pathways: the ventral and dorsal streams. One study demonstrated a double dissociation of visual recognition (face perception camouflaged by shadows) and visuospatial performance (maze learning) in two patients with lesions of the occipitotemporal and occipitoparietal cortex, respectively (Newcombe, Ratcliff, & Damasio, 1987). The specific clinical syndromes produced by occipitotemporal lesions can include visual object agnosia, prosopagnosia (an inability to recognize familiar faces, see below), and achromatopsia, or cortical color blindness. In contrast, syndromes produced by occipitoparietal lesions can include optic ataxia (misreaching), visuospatial neglect, constructional apraxia, gaze apraxia, akinetopsia (an inability to perceive movement), and disorders of spatial cognition. Interestingly, imagery disorders involving descriptions of either objects (especially faces, animals, and colors of objects) or spatial relations (geographic directions) are also dissociable following temporal and parietal lesions, respectively.

The development of functional brain imaging, including PET (positron emission tomography) and fMRI (functional magnetic resonance imaging), has made it possible to map the organization of the human visual cortex with far greater precision than is possible with human lesion studies, and without the confounding influence of compensatory responses to brain injury. This has provided researchers the tools to test hypotheses concerning visual perception in the intact human brain. Haxby, Horwitz, Ungerleider, Maisog, Pietrini, and Grady (1994) used positron emission tomography (PET) to demonstrate that humans, like monkeys, show two visual pathways: a dorsal or occipitoparietal pathway, stretching from V1 through extrastriate to parietal cortex, which is primarily concerned with visuospatial processing, and a ventral or occipitotemporal pathway, stretching from V1 through extrastriate to temporal cortex, which is primarily concerned with object processing (Fig. 44.6). These findings indicate the existence in humans, as in monkeys, of two functionally specialized and anatomically segregated visual processing pathways. They further suggest the extensions of each into the frontal lobe, which plays a role in working memory for objects and their spatial locations (see Chapter 50).

The "Building Blocks" of Object Recognition in Human Cortex Include Early Visual Areas and the Lateral Occipital Complex

Functional neuroimaging studies have revealed a number of distinct visual areas within the ventral and dorsal streams in humans. Many of these areas appear to be equivalent (and perhaps homologous) to specific monkey visual areas, including V1, V2, V3, V3A, V4, and the middle temporal area (MT). As in monkeys, most of these areas have been defined on the basis of retinotopic maps (Sereno et al., 1995). Area MT, a small area with somewhat disorganized retinotopy, instead has been defined on the basis of its distinctive functional properties (strong selectivity for visual motion and direction) in both human neuroimaging and single-unit recordings in macaque.

As described earlier, receptive field size increases, and stimulus requirements become more stringent, as one progresses from V1 through various stages of the ventral stream. For instance, single neurons in macaque IT cortex do not respond to "scrambled" versions of object images (e.g., Fig. 44.5). To test for analogous regions in human visual cortex, Malach et al. (1995) compared responses evoked by a range of images to those evoked by "scrambled" versions of those same images, while attempting to equate several image properties (such as size and contrast). Regions responding more to objects than to scrambled stimuli were interpreted as being involved in object-related processing; collectively these regions were named the lateral occipital complex (LOC) because of their anatomical location (Fig. 44.7).

A number of fMRI findings suggest that the LOC is an important stage in human object processing. First, whereas early visual areas V1–V3 respond similarly to intact, recognizable stimuli and highly scrambled, unrecognizable pictures (scrambled pictures tend to elicit even stronger responses), the LOC *exhibits a high sensitivity to image scrambling*. Second, a fourfold change in visual object size does not affect the LOC activation, although it changes some stimulus properties greatly (e.g., local contrast); responses in the LOC also exhibit some invariance to changes in image position. Third, the LOC responds to multiple visual cues. For example, the shape of an object can be defined by luminance, by texture, or by motion cues alone; objects defined by any one or more of these cues reliably evoke responses in the LOC. Fourth, objects that vary widely in their recognizability (e.g., famous faces and unfamiliar three-dimensional abstract sculptures) produce similar activations in the LOC. Together, these findings indicate that the LOC is an intermediate link in the processing of objects; that is, a stage following low-level processing and preceding object-processing stages that involve memory.

Summary

Evidence from lesion studies and neuroimaging studies confirms the existence in humans, as in macaques, of two visual processing pathways: a ventral stream for object

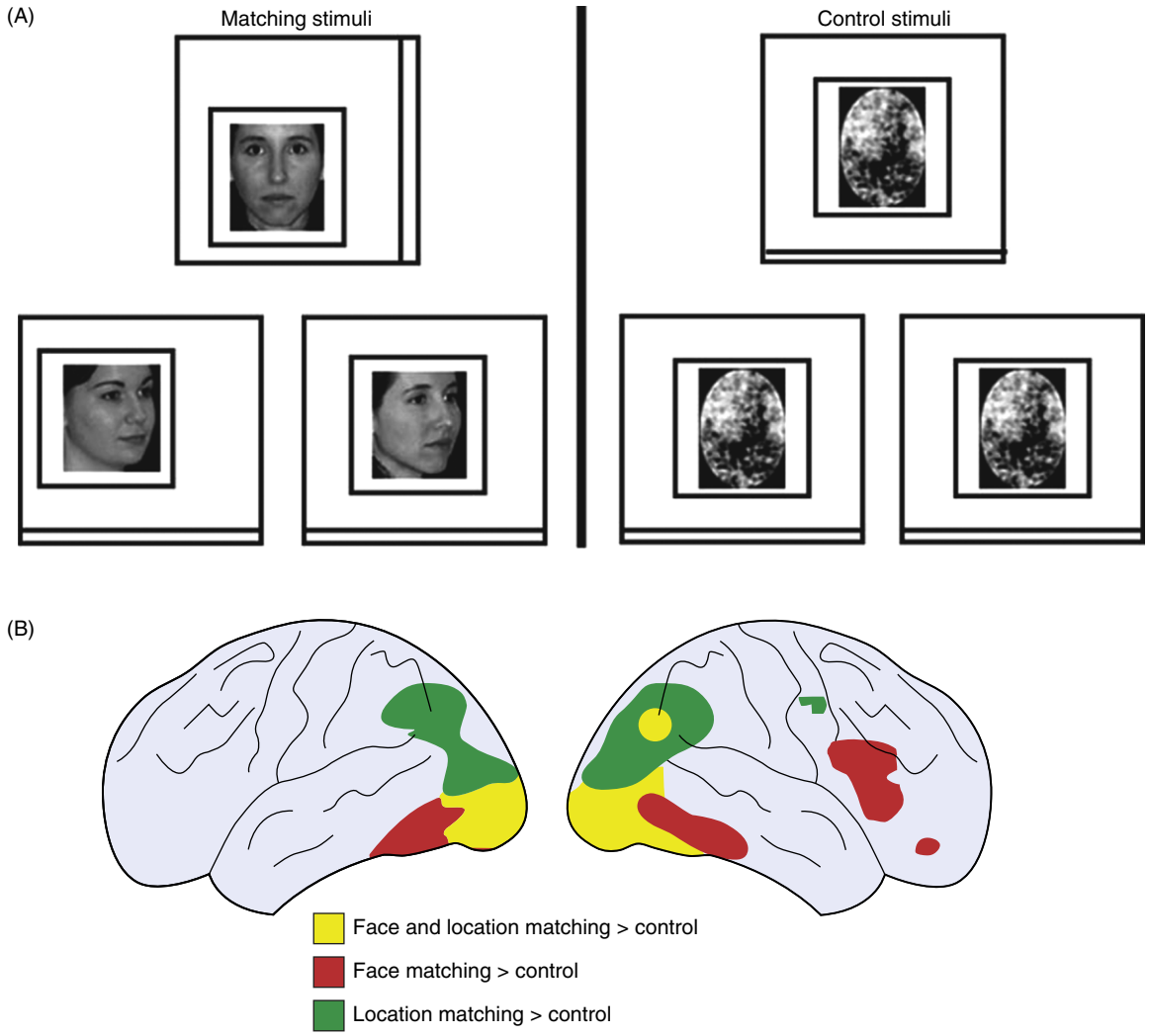


FIGURE 44.6 Ventral and dorsal processing streams in the human cortex, as demonstrated in a PET imaging study of face and location perception. (A) Sample stimuli used during the experiment. (Left) Stimuli for both face and location matching tasks. During face matching trials, subjects had to indicate with a button press which of the two faces at the bottom was the same person shown at the top. In this example, the correct choice is the stimulus on the right. During location matching trials, subjects had to indicate with a button press which of the two small squares at the bottom was in the same location relative to the double line as the small square shown at the top. In this example, the correct choice is the stimulus on the left. (Right) Control stimuli. During control trials, subjects saw an array of three stimuli in which the small squares contained a complex visual image. In these trials, which controlled for both visual stimulation and finger movements, they alternated left and right button presses. (B) Areas shown in red had significantly increased activity during the face matching but not during the location matching task, as compared with activity during the control task. Areas shown in green had significantly increased activity during the location matching but not during the face matching task, as compared with activity during the control task. Areas shown in yellow had significantly increased activity during both face and location matching tasks. *Adapted from Haxby et al. (1994).*

vision and a dorsal stream for spatial vision. Within the ventral stream, early visual areas, including V1–V4, are organized retinotopically and appear to be homologous to areas studied in monkeys. The LOC appears to play an important role at the next stage of processing, where general object shape is processed. Activation of the LOC is highly sensitive to image scrambling, but less sensitive to variations in image size or the visual cues that define the shape of an object. Activation of the LOC, however, is not sensitive to object recognizability, suggesting that it precedes processing stages that involve memory.

PERCEPTION AND RECOGNITION OF SPECIFIC CLASSES OF OBJECTS

Prosopagnosia is a Specific Deficit of Face Recognition

Evidence from neuropsychology and neurophysiology suggests that certain kinds of objects may have special significance to primates. *Prosopagnosia*, a selective deficit in recognizing familiar faces, has been known since at least the end of the nineteenth century. Although patients with

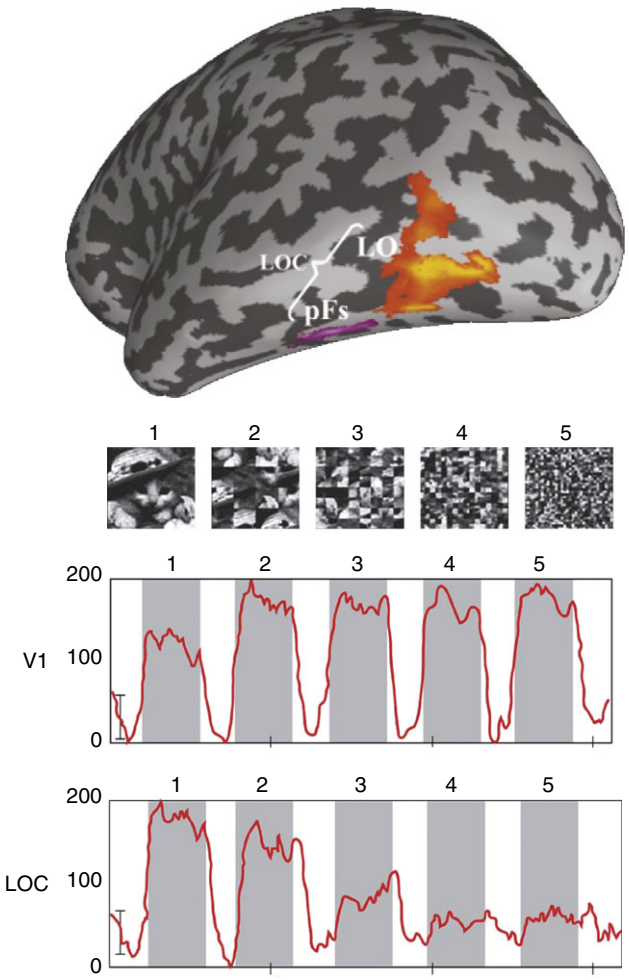


FIGURE 44.7 Shape processing in the lateral occipital cortex (LOC). (Top) Inflated view of the human brain showing the location of the LOC. *Adapted from Malach, Levy, and Hasson (2002).* (Middle) Visual stimuli that were used to investigate the effects of image scrambling on activity in the LOC, as measured by fMRI. Subjects viewed the stimuli and were instructed to covertly name them, even the scrambled images. Stimulus 1 is the original image, and stimuli 2–5 are increasingly scrambled. (Bottom) fMRI time series illustrating the averaged activations in V1 and the LOC to each of the images shown above. Note that the LOC is very sensitive to image scrambling, whereas V1 is not. *Adapted from Grill-Spector et al. (1998).*

this syndrome are aware that faces are faces (i.e., they know the basic stimulus category), they fail to reliably identify (or even achieve a sense of familiarity from) the faces of coworkers, family members, famous persons, and other individuals previously well known to them. Typically, prosopagnosics also have trouble forming memories of new faces, even if other new objects can be learned. Because the voice of a visually unrecognized person usually enables the patient to identify and feel familiar with that person, prosopagnosia (face agnosia) appears to be a specifically visual impairment.

Since the 1980s, a number of cases with prosopagnosia have come to autopsy. The damage common to these lay within the *lingual and fusiform gyri* (Fig. 44.1), ventrally and medially within the cortex at the occipitotemporal junction. In all such cases, this area (or at least the underlying white matter) was damaged bilaterally, and consequently bilateral damage was thought by many to be a necessary precondition for prosopagnosia. Additional cases with prosopagnosia and right cortical damage alone, along with the results of neuroimaging studies (see later), have reopened the debate over the locus of the critical lesion in prosopagnosia.

Several recent reports of developmental prosopagnosia have provided invaluable information concerning the perceptual deficit observed in these patients. Developmental prosopagnosia is characterized by severely impaired face recognition in individuals with no detectable brain damage (i.e., presumably intact visual systems). Many developmental prosopagnosics have normal or relatively good object recognition, indicating that their impairments are not the result of deficits to a unitary visual recognition mechanism. Indeed, extensive testing of particular individuals has revealed especially pure face processing deficits (Duchaine & Nakayama, 2006). In such individuals, the results strongly suggest that faces are recognized by a content-specific, face-processing mechanism. However, it should be noted that neuroimaging studies have demonstrated a variety of neural profiles in developmental prosopagnosia, which is consistent with behavioral studies indicating that, in general, developmental prosopagnosia is a heterogeneous disorder (i.e., in some cases deficits are not restricted to faces).

Neuroimaging and Electrophysiological Studies Reveal Neural Substrates of Prosopagnosia

The existence of prosopagnosia suggests that certain regions within the ventral stream should be very responsive to faces. This is exactly what has been observed in neuroimaging studies. For example, early PET and fMRI studies revealed a region on the fusiform gyrus of the right hemisphere that is significantly more active when subjects view faces, compared to when they view common nonface objects (Fig. 44.8). Kanwisher, McDermott, and Chun (1997) called this area the "fusiform face area" (FFA). The FFA is more active during face matching than location matching, and during the viewing of faces than during the viewing of scrambled faces, letter strings, or textures. Other face-selective regions have since been identified in ventral occipital cortex (the "occipital face area," OFA), the superior temporal sulcus, and the anterior portion of the temporal lobe. Notably, damage to one or more of these areas can lead to deficits in face processing, including prosopagnosia.

Evidence from electrophysiological recordings in patients with implanted electrodes (as part of the preparation for neurosurgery to treat epilepsy) also localizes

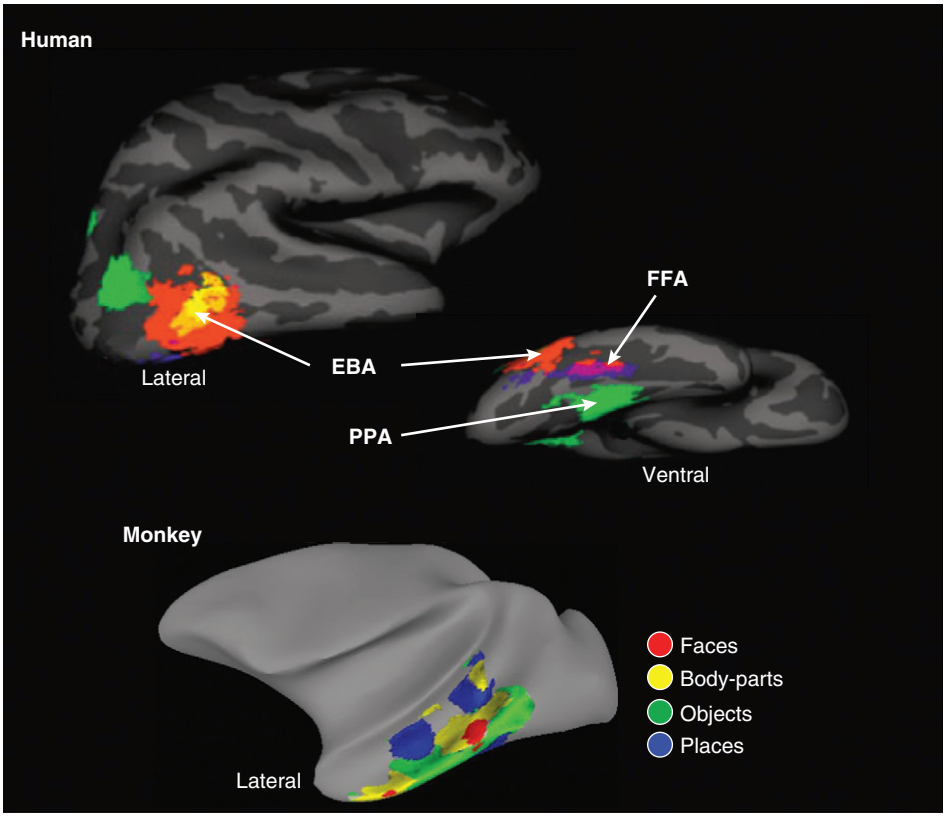


FIGURE 44.8 Category-selective regions in the human and monkey as revealed by fMRI. (Top) Regions in the human brain significantly more activated by faces (fusiform face area, "FFA"), body parts (extrastriate body area, "EBA"), and places (parahippocampal place area, "PPA"). *Adapted from Spiridon, Fischl, and Kanwisher (2006).* (Bottom) Similar regions in the monkey brain. *Adapted from Bell et al. (2009).*

face-selective processing to the ventral occipitotemporal cortex (Allison, Puce, Spencer, & McCarthy, 1999). A large evoked potential, called the *N200* because it occurs roughly 200 ms after stimulus onset, is generated in small regions (i.e., typically at only one electrode within an electrode array) in the ventral occipitotemporal cortex. The N200 amplitude is largest in response to full faces and decreases progressively to eyes, face contours, lips, and noses viewed in isolation. Interestingly, a region just lateral to face-selective N200 sites is more responsive to internal face parts than to faces, and some sites in ventral occipitotemporal cortex appear to be face-part specific. Other electrophysiological approaches have been applied to the study of face processing in humans. Liu, Harris, and Kanwisher (2002) used magnetoencephalography (MEG) to associate two separate evoked potentials (M100 and M170) in the occipitotemporal cortex to different stages of face processing: face detection and face identification.

Ventral Pathway Contains Regions Specialized for Other Visual Categories

Subsequent neuroimaging studies have identified regions selective for other visual categories. Medial to the fusiform gyrus is an area within parahippocampal cortex that is significantly more active when subjects view complex scenes (such as rooms, landscapes, landmarks, and city streets) compared to when they view objects, faces, or other kinds of visual stimuli. This region has been named the parahippocampal place area (PPA) (Fig. 44.8) (Epstein & Kanwisher, 1998) and damage to the lingual/parahippocampal region may result in landmark agnosia. Along the lateral occipitotemporal cortex (near the middle occipital gyrus), an area selectively activated by images of human body-parts (such as an isolated foot or arm) relative to viewing objects has been identified (Downing, Jiang, Shuman, & Kanwisher, 2001; Orlov, Makin, & Zohary, 2010). This region has been called the extrastriate body area (EBA) (Fig. 44.8). An additional body-part selective region has been identified in the fusiform gyrus (the "fusiform body area," FBA) (Schwarzlose, Baker, & Kanwisher, 2005).

Studies of brain lesions in patients suggest that there is a specialized region for the processing of *written words* in the brain. Patients with *alexia* cannot read despite intact visual capabilities, and despite their ability to understand spoken language and to write. Alexia can therefore be conceptualized as essentially a visual agnosia for verbal material. Accordingly, an area has been discovered near the fusiform gyrus in the occipitotemporal and inferior occipital sulci (Cohen et al., 2000). This "visual word form

area" (VWFA) is selectively activated by letters, words, and to a lesser degree, nonsense letterstrings. It is found almost exclusively in the left hemisphere, the same hemisphere as Broca's and Wernicke's areas—regions intimately involved in speech and language comprehension. Importantly, because we are not born with knowledge about written words, the existence of VWFA indicates that at least some category-specific cortex must arise as a result of experience.

Collectively, this evidence suggests that the human visual system is organized such that some object categories are represented in a relatively specialized fashion: faces, places, body parts, and words appear to have dedicated neural mechanisms for their representation. Similar specialization has been observed in the macaque brain using fMRI. Specifically, presenting images of faces (human or monkey), body-parts (human or monkey), objects, or places activates discrete areas within the inferior and ventral temporal cortex (Fig. 44.8) (Tsao, Freiwald, Knutsen, Mandeville, & Tootell, 2003; Pinsk, DeSimone, Moore, Gross, & Kastner, 2005; Pinsk et al., 2009; Bell, Hadj-Bouziane, Frihauf, Tootell, & Ungerleider, 2009). As described above, neuronal recordings from these regions have revealed disproportionately high concentrations of neurons selective for these categories (Tsao et al., 2006; Bell et al., 2011). Recent developments in fMRI methods have further demonstrated that these regions form interconnected networks specialized for processing stimuli from different visual categories (see Box 44.3).

Category-Selective Regions Linked to Conscious Perception

The existence of specialized regions in the temporal cortex clearly argues for a strong role in visual perception—but what precisely is this role? One possibility, for which there is mounting evidence, is that these areas are not simply responding to the visual features that comprise stimuli from a given category, but that they are involved in the *conscious perception* of these stimuli. In one neuroimaging study, the differential sensitivity of the ventral cortex to different visual stimuli was explored in a *binocular rivalry* paradigm in which different stimuli (faces and houses) were presented to the right and left eyes (Tong, Nakayama, Vaughan, & Kanwisher, 1998) (Fig. 44.9). During binocular rivalry, the resulting percept is not a combination of the two stimuli. Instead, subjects report the actual percept to alternate between the two stimuli, every few seconds—even though the dual retinal stimulation is kept constant. As expected, subjects' perceptions fluctuated between

BOX 44.3

LINKING CATEGORY-SELECTIVE REGIONS INTO SPECIALIZED NETWORKS

The application of fMRI to the study of object perception has resulted in the identification of a number of regions within the occipitotemporal cortex sensitive to particular categories of visual stimuli, including faces, body parts, and places. In virtually all cases, multiple regions for each category have been observed; suggesting the existence of separate *networks* specialized for the processing of specific categories. Until recently, demonstrating connectivity between individual nodes within these putative networks would have required classical anatomical tracer studies. However, recent developments in MRI analytic techniques have provided researchers with the opportunity to study connectivity in the intact brain.

Resting state or functional connectivity (FC) analysis, for example, measures the degree to which fMRI activation in one area is correlated with activation in another, and can be used to identify not only possible connections between areas but also offer insight as to the nature of the interactions (see Deco, Jirsa, & McIntosh, 2011). Zhu, Zhang, Luo, Dilks, and Liu (2011) used FC to show how the correlation between activation in FFA and OFA was greater during face recognition tasks as compared to object recognition tasks. Nummenmaa, Passamonti, Rowe, Engell, and Calder (2009) used a similar approach to show how face-processing networks interact with the oculomotor network in the frontal cortex during direct shifts in attention induced by changes in gaze direction.

Diffusion tensor imaging (DTI), which can be used to identify white matter tracts (see Johansen-Berg & Rushworth, 2009), has been used to show how face perception ability relates to the viability of connections between the various face-selective regions in human temporal cortex. Patients suffering from congenital prosopagnosia show reduced white matter volume in the fiber tracts travelling through the fusiform gyrus to the anterior temporal cortex (Thomas, Avidan, Humphreys, Jung, Gao, & Behrmann, 2009). Finally, in the nonhuman primate, fMRI can be used in conjunction with microstimulation to visualize connections between functionally related regions. Moeller, Freiwald, and Tsao (2008) applied microstimulation to fMRI-identified face patches to show that the nonhuman primate brain contains an interconnected network of regions dedicated to face processing.

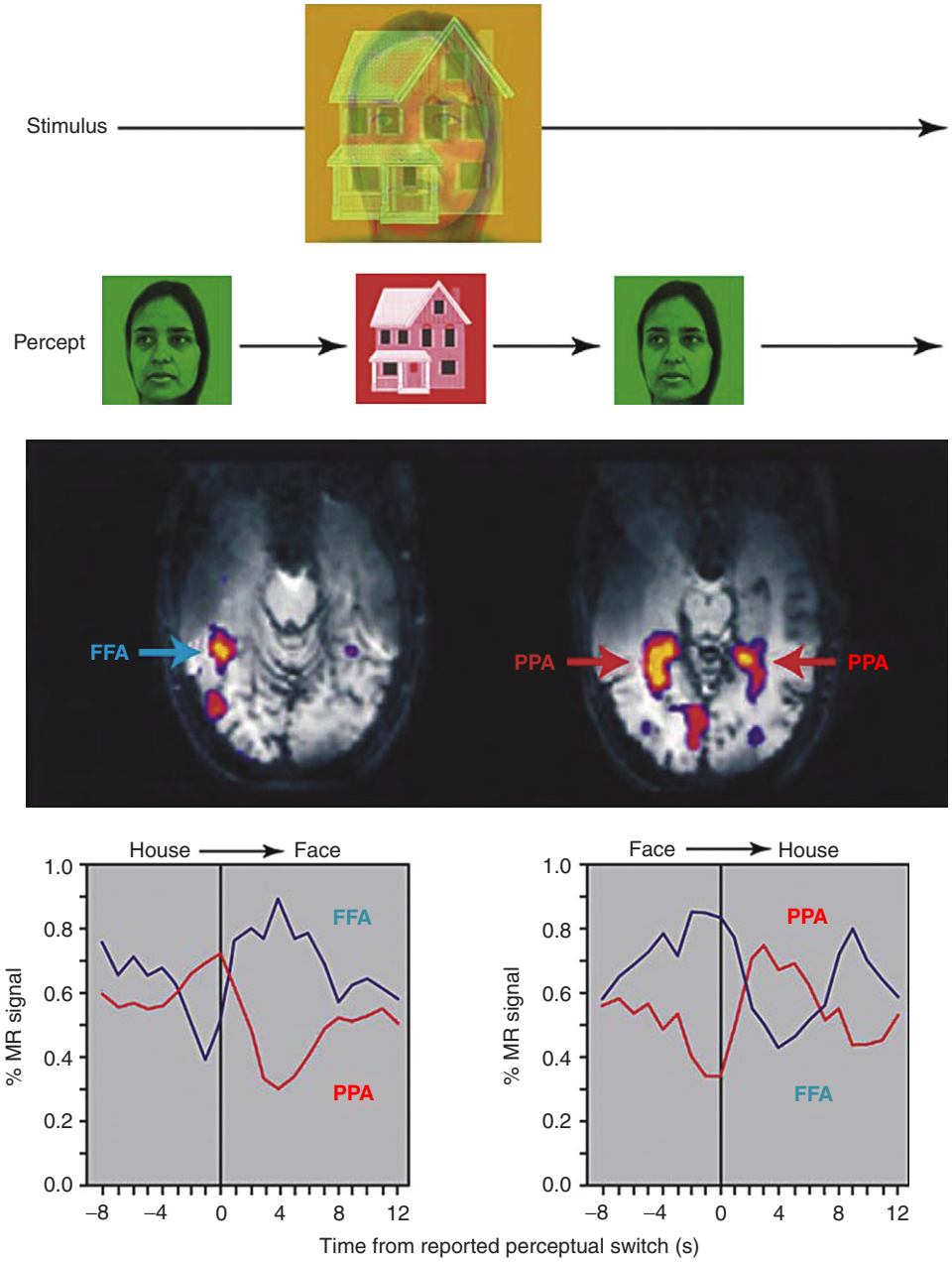


FIGURE 44.9 Design and results from an fMRI experiment on binocular rivalry. (Top) The illustration shows the ambiguous face/house stimulus used to produce rivalry. When viewed through red and green filter glasses, only the face could be seen through one eye and only the house through the other eye. This arrangement led to binocular rivalry, with a face percept alternating with a house percept every few seconds. (Middle) Two adjacent axial slices through the brain of a single subject showing, in the left slice, the area on the fusiform gyrus that responded more to faces than houses (FFA) and, in the right slice, the area on the parahippocampal gyrus that responded more to houses than to faces (PPA). In the slices shown, left is the right hemisphere and right is the left hemisphere. The scale at right indicates (Bottom) fMRI time series showing FFA and PPA activity for a single subject illustrating that the activity in these two regions correlated with the subject's visual percept. FFA, fusiform face area; PPA, parahippocampal place area. *Adapted from Tong et al. (1998).*

faces and houses. Critically, while changes from house to face percepts were accompanied by increased activation in the right FFA, changes from face to house percepts led to decreased activation in the FFA.

Summerfield, Egner, Mangels, and Hirsch (2006) demonstrated that FFA responds even during incorrect perceptual judgments concerning faces. On each trial, subjects were presented with either a degraded face or house and were required to report which of the two categories was presented. Predictably, PPA was activated when subjects correctly indicated seeing a house, and FFA and OFA were activated when subjects correctly indicated seeing a face. However, FFA activation was also increased when subjects *incorrectly* reported seeing a face when, in fact, a house was presented. This suggests that activation in FFA is related more to the *percept* than to the physical stimulus itself.

Artificially influencing these regions can directly impact perception. Afraz et al. (2006) trained monkeys to perform a categorization task where they were presented with degraded images of faces and non-face objects. By applying electrical microstimulation to small clusters of neurons located in inferior temporal cortex (likely in the vicinity of the face-selective patches, see above), they were able to bias the monkey's decision toward faces irrespective of the stimulus presented.

Alternative Proposals to Category-Specific Representations

It has been proposed that the FFA is specialized not for the processing of faces *per se* but rather for the processing of objects with which one has acquired expertise (Gauthier, Williams, Tarr, & Tanaka 1998; Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999). The idea that expertise underlies responses in ventral temporal cortex raises an important distinction between two interpretations. One interpretation suggests that activation is based on visual features (e.g., the visual features composing a face, such as eyes, nose, and mouth). Another interpretation emphasizes the *processes* associated with perceiving or recognizing objects. In this latter view, the ventral stream contains areas that are best suited for different computations, which may be recruited as one acquires expertise with specific types of visual objects. This might involve encoding subtle differences between visually similar objects or, conversely, finding underlying similarities between visually dissimilar objects. In particular, the expertise view may help explain why some patients with prosopagnosia may also have associated deficits in areas of acquired expertise, such as a show dog expert who has lost the ability to differentiate breeds or a farmer who has lost the ability to recognize the cows in his herd. According to another alternative framework, the functional architecture of the ventral stream is viewed not as being composed of category-specific regions but instead as containing a distributed representation of information about object form. Consistent with this hypothesis, several fMRI studies have shown how different categories can be associated with individual patterns of response across the ventral temporal cortex in a highly consistent spatial arrangement across subjects (Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999; Haxby, Gobbini, Furey, Ishai, Schouten, & Pietrini, 2001). Moreover, it is further possible to accurately identify which stimulus was presented to a subject based on the elicited pattern of fMRI activation (e.g., Kay, Naselaris, Prenger, & Gallant, 2008). In this view, the representation of an object is not restricted to a region that responds maximally to that object, but rather is distributed across a broader expanse of cortex. Within this distributed system, however, the representation of faces would be less extensive as compared to the representation of nonface objects, helping explain why it is possible for lesions to produce prosopagnosia while leaving the recognition of nonface objects largely intact.

Object Knowledge Is Stored in a Distributed Network of Cortical Areas

Although the distinction between perception and memory might seem sharp, in reality these two processes are intimately connected. Indeed, each time one identifies an object; one implicitly retrieves stored knowledge about that object. In other words, identifying an object necessarily implies knowing what that object is. Further, it is generally believed that the same neural tissue devoted to object processing also participates in the storage of the long-term representations of objects. For example, when a subject mentally imagines faces or houses (requiring the retrieval of stored object representations from long-term memory), the same patterns of activation across ventral temporal cortex are seen as those observed when actually viewing faces or houses. Object knowledge implies not only familiarity with the shape of an object, but also other attributes, such as its typical color and function. For example, how does one know that bulldozers are usually yellow and are used for digging? One neuroimaging study that examined this question asked subjects to name either the associated color or the associated action in response to the presentation of black and white line drawings of objects (Martin, Haxby, Lalonde, Wiggs, & Ungerleider, 1995). By associated action the investigators meant the typical motion associated with an object, such as the motion produced by the arm of the bulldozer digging. Results showed that the generation of color names associated with objects selectively activated a region in the ventral occipital cortex just in front of the area involved in the perception of color, whereas the generation of action names activated a region dorsally in the middle temporal gyrus just in front of the area involved in the perception of motion. These findings suggest that information about the different attributes of an object is not stored in a unified fashion in any single area of the cortex. Rather, object knowledge seems to be stored in a distributed fashion such that information about specific attributes is stored close to the specific regions of the cortex that mediate the perception of those attributes. These neuroimaging results provide a basis for understanding category-specific disorders of knowledge resulting from patterns of focal brain damage.

This type of deficit, seen most commonly for animals and tools, involves a selective difficulty in naming and retrieving information about objects from a single category. If the distinction between different categories of objects, such as animals and tools, is dependent on access to information about different types of attributes, and if such attributes (e.g., form, color, motion and object-use associated motor movements) are stored in separate regions of the brain, then one may find a category-specific deficit depending on the site of the lesion. Indeed, consistent with neuroimaging data, it has been found that the locations of lesions in patients with impaired recognition and impaired naming of drawings of animals included the medial aspects of the occipital lobe, nearly the same region active during animal naming. In contrast, the locations of lesions in patients with impaired recognition and

impaired naming of drawings of tools all included the middle temporal gyrus, nearly the same region active during tool naming (Tranel, Damasio, & Damasio, 1997). It is important to note that the impairment in patients with category-specific deficits is not limited to visual recognition. These deficits occur when knowledge is probed visually or verbally, and therefore they reflect disorders of stored knowledge and not simply disorders of perception.

Summary

Prosopagnosia, a selective deficit in recognizing faces, is associated with damage to the occipitotemporal cortex, including the fusiform gyrus. Consistent with this impairment, neuroimaging and electrophysiological studies have demonstrated that the fusiform gyrus responds selectively to faces and that this activation is linked to face perception. Although these findings suggest the existence of a specialized region for processing faces, an alternative interpretation is that the fusiform gyrus performs specialized computations associated with perceiving certain objects, including but not limited to faces. Such computations would presumably apply to many objects with which one has expertise. Moreover, the existence of a region in the left occipitotemporal cortex selectively activated by letter strings and words demonstrates the role of experience in shaping the functional architecture of the object recognition pathway.

OVERALL SUMMARY

Early studies of brain damage in humans and lesions in monkeys pointed to the crucial role of tissue in the ventral parts of the temporal cortex for object perception and recognition. After many decades of intensive research, one can now say that the ventral portions of the temporal cortex form the last stations of an occipitotemporal ventral pathway, beginning in the primary visual cortex and progressing through multiple visual areas beyond it. In macaque monkeys, important regions for object perception and recognition include areas TEO and TE of the inferior temporal cortex. In humans, homologous circuits appear to exist in both occipital and temporal regions, including the lateral occipital complex, the fusiform gyrus, and nearby regions in the ventral temporal cortex.

Single-unit recording studies in monkeys show a progressive increase in the complexity of object features needed to "trigger" neurons at progressively more anterior stations in the occipitotemporal pathway. Many TE neurons show preservation of their selectivity over transformations that change the physical stimulus, such as retinal position, size, distance in depth, and degree of ambient illumination. At the highest levels of the pathway, within area TE, some neurons are preferentially responsive to faces and other biologically significant stimuli.

More recently, neuroimaging and electrophysiological studies in humans with implanted electrodes have also shed light on the mechanisms by which the primate brain represents objects. Descriptions of prosopagnosia, a selective deficit of face recognition that is associated with damage to ventral occipitotemporal cortex, foreshadowed discoveries of cortical zones that are activated preferentially by face stimuli. At the same time, selectivity for letter strings and even for learned arbitrary stimuli has suggested that experience also influences activity in the ventral pathway and, in the long term, may tune representations. Finally, more general knowledge about objects (what they are used for and how to manipulate them) appears to be stored across a distributed network of sensory and motor cortical areas.

References

Afraz, S. R., Kiani, R., & Esteky, H. (2006). Microstimulation of inferotemporal cortex influences face categorization. *Nature, 442,* 692–695.

Allison, T., Puce, A., Spencer, D. D., & McCarthy, G. (1999). Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cerebral Cortex, 9,* 415–430.

Bell, A. H., Hadj-Bouziane, F., Frihauf, J. B., Tootell, R. B., & Ungerleider, L. G. (2009). Object representations in the temporal cortex of monkeys and humans as revealed by functional magnetic resonance imaging. *Journal of Neurophysiology, 101,* 688–700.

Bell, A. H., Malecek, N. J., Morin, E. L., Hadj-Bouziane, F., Tootell, R. B. H., & Ungerleider, L. G. (1999). Relationship between fMRI-identified regions and neuronal category-selectivity. *Journal of Neuroscience, 31,* 12229–12240.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review, 94,* 115–147.

Biederman, I. (1990). Higher-level vision. In D. H. Osherson, S. M. Kosslyn, & J. M. Hollerbach (Eds.), *Visual cognition and action* (pp. 41–72). Cambridge, MA: MIT Press.

Brincat, S. L., & Connor, C. E. (2006). Dynamic shape synthesis in posterior inferotemporal cortex. *Neuron, 49,* 17–24.

Brown, S., & Schafer, E. A. (1888). An investigation into the functions of the occipital and temporal lobes of the monkey's brain. *Philosophical Transactions of the Royal Society of London, 179,* 303–327.

Bruce, C., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *Journal of Neurophysiology, 46,* 369–384.

Bulthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences of the United States of America, 89,* 60–64.

Bulthoff, H. H., Edelman, S. Y., & Tarr, M. J. (1995). How are three-dimensional objects represented in the brain? *Cerebral Cortex, 5,* 247–260.

Cohen, L., Dehaene, S., Naccache, L., Lehericy, S., Dehaene-Lambertz, G., Henaff, M. A., & Michel, F. (2000). The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain: A Journal of Neurology, 123*(Pt 2), 291–307.

Deco, G., Jirsa, V. K., & McIntosh, A. R. (2011). Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nature Reviews Neuroscience, 12,* 43–56.

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science, 293,* 2470–2473. New York.

Duchaine, B. C., & Nakayama, K. (2006). Developmental prosopagnosia: a window to content-specific face processing. *Current Opinion in Neurobiology, 16,* 166–173.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392,* 598–601.

Erickson, C. A., Jagadeesh, B., & Desimone, R. (2000). Clustering of perirhinal neurons with similar properties following visual experience in adult monkeys. *Nature Neuroscience, 3,* 1143–1148.

Farah, M. J. (2004). *Visual agnosia* (2nd ed.). Cambridge: MIT Press.

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training "greeble" experts: a framework for studying expert object recognition processes. *Vision Research, 38,* 2401–2428.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform "face area" increases with expertise in recognizing novel objects. *Nature Neuroscience, 2,* 568–573.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences, 15,* 20–25.

Grill-Spector, K., Kushnir, T., Hendler, T., Edelman, S., Itzchak, Y., & Malach, R. (1998). A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Human Brain Mapping, 6,* 316–328.

Gross, C. G. (2002). Genealogy of the "grandmother cell." *Neuroscientist, 8,* 512–518.

Gross, C. G., Bender, D. B., & Rocha-Miranda, C. E. (1969). Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science, 166,* 1303–1306. New York.

Gross, C. G., Rodman, H. R., Gochin, P. M., & Columbo, M. W. (1993). Inferior temporal cortex as a pattern recognition device. In E. Baum (Ed.), *Computational learning and cognition* (pp. 44–73). Philadelphia: Society for Industrial and Applied Mathematics.

Haxby, J. V., Horwitz, B., Ungerleider, L. G., Maisog, J. M., Pietrini, P., & Grady, C. L. (1994). The functional organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations. *Journal of Neuroscience, 14,* 6336–6353.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293,* 2425–2430. New York.

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences of the United States of America, 96,* 9379–9384.

Johansen-Berg, H., & Rushworth, M. F. (2009). Using diffusion imaging to study human connectional anatomy. *Annual Review of Neuroscience, 32,* 75–94.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17,* 4302–4311.

Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature, 452,* 352–355.

Kobatake, E., Wang, G., & Tanaka, K. (1998). Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *Journal of Neurophysiology, 80,* 324–330.

Koffka, K. (1935). Principles of gestalt psychology. *Psychological Bulletin, 19,* 531–585.

Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature Reviews Neuroscience, 12,* 217–230.

Li, N., & DiCarlo, J. J. (2008). Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *Science, 321,* 1502–1507. New York.

Li, N., & DiCarlo, J. J. (2010). Unsupervised natural visual experience rapidly reshapes size-invariant object representation in inferior temporal cortex. *Neuron, 67,* 1062–1075.

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: an MEG study. *Nature Neuroscience, 5,* 910–916.

Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology, 5,* 552–563.

Malach, R., Levy, I., & Hasson, U. (2002). The topography of high-order human object areas. *Trends in Cognitive Sciences, 6,* 176–184.

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., & Tootell, R. B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences of the United States of America, 92,* 8135–8139.

Martin, A., Haxby, J. V., Lalonde, F. M., Wiggs, C. L., & Ungerleider, L. G. (1995). Discrete cortical regions associated with knowledge of color and knowledge of action. *Science, 270,* 102–105. New York.

Messinger, A., Squire, L. R., Zola, S. M., & Albright, T. D. (2001). Neuronal representations of stimulus associations develop in the temporal lobe during learning. *Proceedings of the National Academy of Sciences of the United States of America, 98,* 12239–12244.

Miller, E. K., & Desimone, R. (1994). Parallel neuronal mechanisms for short-term memory. *Science, 263,* 520–522. New York.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences, 6,* 415–417.

Moeller, S., Freiwald, W. A., & Tsao, D. Y. (2008). Patches with links: a unified system for processing faces in the macaque temporal lobe. *Science, 320,* 1355–1359. New York.

Munk, H. (1881). On the functions of the cortex. In: Uber die funktionen der Grosshirnrinde, 3te Millteilung. In G. Von Bonin (Ed.), *The cerebral cortex* (pp. 97–117). Springfield: Thomas.

Newcombe, F., Ratcliff, G., & Damasio, H. (1987). Dissociable visual and spatial impairments following right posterior cerebral lesions: clinical, neuropsychological and anatomical evidence. *Neuropsychologia, 25,* 149–161.

Nummenmaa, L., Passamonti, L., Rowe, J., Engell, A. D., & Calder, A. J. (2009). Connectivity analysis reveals a cortical network for eye gaze perception. *Cerebral Cortex, 20,* 1780–1787.

Op de Beeck, H. P., Deutsch, J. A., Vanduffel, W., Kanwisher, N. G., & DiCarlo, J. J. (2008). A stable topography of selectivity for unfamiliar shape classes in monkey inferior temporal cortex. *Cerebral Cortex, 18,* 1676–1694.

Orlov, T., Makin, T. R., & Zohary, E. (2010). Topographic representation of the human body in the occipitotemporal cortex. *Neuron, 68,* 586–600.

Pasupathy, A., & Connor, C. E. (2002). Population coding of shape in area V4. *Nature Neuroscience, 5,* 1332–1338.

Pinsk, M. A., DeSimone, K., Moore, T., Gross, C. G., & Kastner, S. (2005). Representations of faces and body parts in macaque temporal cortex: a functional MRI study. *Proceedings of the National Academy of Sciences of the United States of America, 102,* 6996–7001.

Pinsk, M. A., Arcaro, M., Weiner, K. S., Kalkus, J. F., Inati, S. J., Gross, C. G., & Kastner, S. (2009). Neural representations of faces and body parts in macaque and human cortex: a comparative FMRI study. *Journal of Neurophysiology, 101,* 2581–2600.

Rubens, A. B., & Benson, D. F. (1971). Associative visual agnosia. *Archives of Neurology, 24,* 305–316.

Sakai, K., & Miyashita, Y. (1991). Neural organization for the long-term memory of paired associates. *Nature, 354,* 152–155.

Schwarzlose, R. F., Baker, C. I., & Kanwisher, N. (2005). Separate face and body selectivity on the fusiform gyrus. *Journal of Neuroscience, 25,* 11055–11059.

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., Rosen, B. R., & Tootell, R. B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science, 268,* 803–804.

Spiridon, M., Fischl, B., & Kanwisher, N. (2006). Location and spatial profile of category-specific regions in human extrastriate cortex. *Human Brain Mapping, 27,* 77–89.

Summerfield, C., Egner, T., Mangels, J., & Hirsch, J. (2006). Mistaking a house for a face: neural correlates of misperception in healthy humans. *Cerebral Cortex, 16,* 500–508.

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience, 19,* 109–139.

Tanaka, K., Saito, H., Fukada, Y., & Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *Journal of Neurophysiology, 66,* 170–189.

Thomas, C., Avidan, G., Humphreys, K., Jung, K. J., Gao, F., & Behrmann, M. (2009). Reduced structural connectivity in ventral visual cortex in congenital prosopagnosia. *Nature Neuroscience, 12,* 29–31.

Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron, 21,* 753–759.

Tranel, D., Damasio, H., & Damasio, A. R. (1997). A neural basis for the retrieval of conceptual knowledge. *Neuropsychologia, 35,* 1319–1327.

Tsao, D. Y., Freiwald, W. A., Tootell, R. B., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science, 311,* 670–674. New York.

Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B., & Tootell, R. B. (2003). Faces and objects in macaque cerebral cortex. *Nature Neuroscience, 6,* 989–995.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. G. Ingle (Ed.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.

Zhu, Q., Zhang, J., Luo, Y. L., Dilks, D. D., & Liu, J. (2011). Resting-state neural activity across face-selective cortical regions is behaviorally relevant. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 31,* 10323–10330.

Suggested Readings

Bar, M. (2004). Visual objects in context. *Nature Review in Neuroscience, 5,* 617–629.

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Review in Neuroscience, 6,* 641–651.

Damasio, A. R. (1990). Category-related recognition deficits as a clue to the neural substrates of knowledge. *Trends in Neuroscience, 13,* 95–98.

Farah, M. J. (2004). *Visual agnosia* (2nd ed.). Cambridge: MIT Press.

Forde, E. M. E., & Humphreys, G. W. (1999). Category-specific recognition impairments: a review of important case studies and influential theories. *Aphasiology, 13,* 169–193.

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review in Neuroscience, 19,* 577–621.

Martin, A., Ungerleider, L. G., & Haxby, J. V. (2000). Category specificity and the brain: The sensory/motor model of semantic representations of objects. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 1023–1036). Cambridge, MA: MIT Press.

Sereno, M. I., & Tootell, B. H. (2005). From monkeys to humans: What do we now know about brain homologies? *Current Opinion in Neurobiology, 15,* 135–144.

Ullman, S. (1996). *High-level vision*. Cambridge, MA: MIT Press.

Ungerleider, L. G., & Pasternak, T. (2004). Ventral and dorsal cortical processing streams. In L. M. Chalupa & J. S. Werner (Eds.), *The visual neurosciences* (pp. 541–562). Cambridge, MA: MIT Press.

Andrew H. Bell, Luiz Pessoa, Roger B.H. Tootell, and Leslie G. Ungerleider

CHAPTER

45

Spatial Cognition

As we move through the world, new visual, auditory, vestibular, and somatosensory inputs continuously impinge on the brain. Given such constantly changing input, it is remarkable how easily we are able to keep track of where things are. We can reach for an object, look at it, or kick it without making a conscious effort to assess its location. We understand the arrangement of parts in an object. We know automatically where we are in the environment and how to get from where we are to where we want to go. How is it that we can achieve the effortless mastery of space that underlies these abilities? This chapter outlines the contributions of brain areas in the parietal, frontal, and temporal lobes to the generation of actions in space, the understanding of spatial relations and navigation in the world.

NEURAL SYSTEMS FOR SPATIAL COGNITION

Dorsal Stream Visual Areas Send Spatial Information to Parietal Cortex

Within the cerebral hemispheres, visual information is processed serially by a succession of areas beginning with the primary visual cortex of the occipital lobe. Within this hierarchical system, there are two parallel chains, or streams, of areas: the ventral stream, leading downward into the temporal lobe, and the dorsal stream, leading forward into the parietal lobe (Ungerleider & Mishkin, 1982) (see Chapter 44). Although the two streams are interconnected to some degree, it is a fair first approximation to describe them as separate parallel systems. Areas of the ventral stream play a critical role in the recognition of visual patterns, including faces, whereas areas within the dorsal stream contribute selectively to conscious spatial awareness and to the spatial guidance of actions, such as reaching and grasping. Dorsal stream areas have several distinctive functional characteristics, including containing a comparatively extensive representation of the peripheral visual field and being specialized for the detection and analysis of moving visual images. These traits would be expected in any system processing visual information for use in spatial awareness and in the visual guidance of behavior.

Although visual input is important for spatial operations, awareness of space is more than just a visual function. It is possible to apprehend the shape of an object and know where it is, regardless of whether it is seen or sensed through touch. Accordingly, spatial awareness, considered as a general phenomenon, depends not simply on visual areas of the dorsal stream but rather on the higher order cortical areas to which they send their output. The transition from areas serving purely visual functions to those mediating generalized spatial awareness is gradual, but has been accomplished by the time the dorsal stream reaches its termination in the association cortex of the parietal lobe.

Three Spatial Processing Streams Emanate from Parietal Cortex

Numerous association areas in the cerebral hemisphere mediate cognitive functions that depend in some way on the use of spatial information originating in parietal cortex. These areas occupy a continuous region of the cerebral hemisphere, encompassing large parts of the frontal, cingulate, temporal, parahippocampal, and insular cortices. They are connected to one another anatomically and to the parietal cortex by a parallel distributed pathway through which signals shuttle back and forth (Goldman-Rakic, 1988). Within this system, there at least three systems of areas with distinguishable functions (Kravitz, Saleem,

© 2013 Elsevier Inc. All rights reserved.

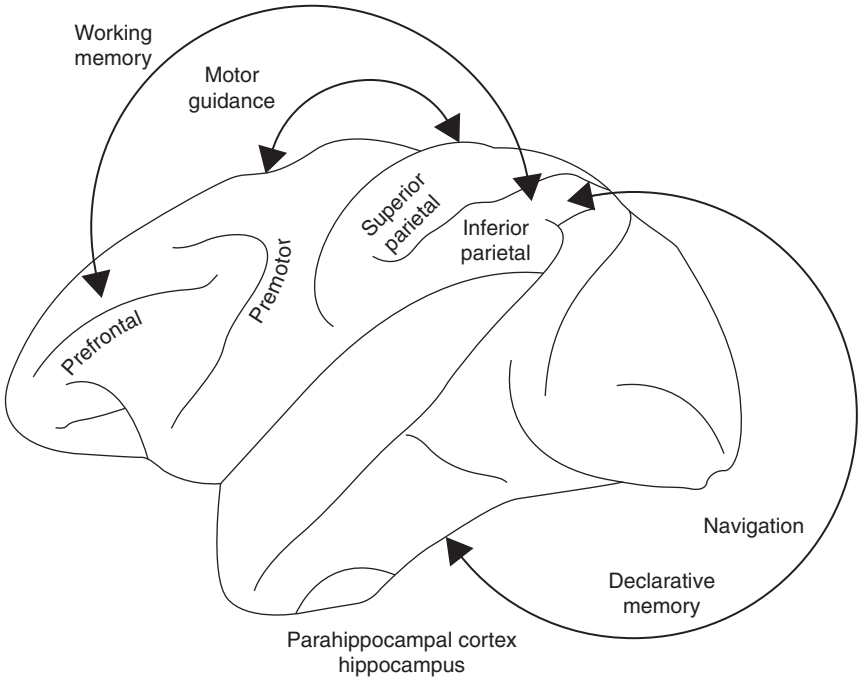


FIGURE 45.1 Lateral view of the left cerebral hemisphere of a rhesus monkey. Pathways emanating from parietal cortex link it to other cortical regions with functions that depend on spatial information. *Brain drawing, courtesy of Laboratory of Neuropsychology, NIMH.*

Baker, & Mishkin, 2011) (Fig. 45.1). The premotor areas of the frontal lobe mediate the spatial guidance of motor behavior in the workspace around the body. The dorsolateral prefrontal cortex mediates spatial functions of higher order including spatial working memory. A system of areas spanning the cingulate gyrus, the medial temporal lobe, and the hippocampus mediates the ability to navigate in the environment and spatial episodic memory. The following sections describe the spatial functions of the parietal cortex and of the far-flung system of areas to which it projects.

PARIETAL CORTEX

The Parietal Cortex Contributes to Spatial Perception and Attention

The parietal lobe is divided into superior and inferior parietal lobules. Within each lobule, several subdivisions are distinguished by anatomical and functional properties (Colby & Duhamel, 1991). In both humans and monkeys, the superior parietal lobule serves functions related primarily to somesthesis, or tactile perception, whereas the inferior parietal lobule serves functions related primarily to visuospatial cognition (Fig. 45.1). The intraparietal sulcus runs in between the superior and the inferior lobules. Within the sulcus there are several functionally distinct areas with independent representations of space. In monkeys, these include the medial intraparietal area (MIP), which is important for reaching behavior, the lateral intraparietal area (LIP), which is important for spatial attention, and the ventral intraparietal area (VIP), where visual and somatosensory representations of space are brought together. Functional imaging studies indicate that there are probably equivalent areas in humans although the detailed pattern of correspondence remains to be worked out (Grefkes & Fink, 2005).

Injury to the Human Parietal Cortex Causes Impairments of Spatial Function

The neurologist Balint in 1909 first described a group of behavioral impairments specifically associated with bilateral damage to the parietal lobe. Balint's syndrome includes difficulty in executing eye movements to engage visual targets (ocular apraxia), an inability to perceive multiple objects at the same time (simultanagnosia), and inaccuracy in reaching for visual targets (optic ataxia). Many other impairments are now known to arise from damage to parietal cortex. These include constructional apraxia, an inability to copy, through drawing or physical manipulation, the spatial pattern in which things are arranged, and hemispatial neglect, a profound unawareness of the half of space opposite the injured hemipshere. The following sections describe in detail the symptoms of optic ataxia, constructional apraxia, and neglect.

Optic Ataxia: An Impairment of Visually Guided Reaching

After damage to the parietal cortex, some patients experience difficulty in making visually guided arm movements. This condition is referred to as misreaching, or

optic ataxia (Perenin & Vighetto, 1988). Patients with optic ataxia experience difficulty in real-life situations requiring them to reach accurately under visual guidance. For example, a patient, when cooking with a frying pan, may see the handle perfectly well and yet be unable to reach accurately for it without looking directly at it (Rondot, de Recondo, & Dumas, 1977). Neuropsychologists have been able to characterize the reaching deficits of such patients in considerable detail by testing them in controlled situations.

In the test illustrated in Figure 45.2, patients were required to reach out and insert one hand into a slot in the center of a disk held by the experimenter. The disk could be held in either the right or the left visual hemifield, while the patient looked straight ahead, and the patient could be asked to reach with either the right or the left hand. To perform the task, the patient had to direct the hand toward the center of the slot and orient the hand correctly so that it would pass into it. Pictures in the top row of Figure 45.2 show a patient with right parietal lobe damage reaching with the left hand. When the disk was positioned in the right (ipsilesional) visual field, the patient was able to reach accurately (Fig. 45.2A). And yet, when the disk was in the left (contralesional) visual field, the patient committed errors of hand orientation (Fig. 45.2B) and of reaching direction (Fig. 45.2C). The lower row of Figure 45.2 shows reaching by a patient with damage to the left parietal lobe. When using the left hand in the left (ipsilesional) visual field, this patient was also able to reach accurately (Fig. 45.2D). However, when using the right hand in the right (contralesional) visual field, the patient committed errors of hand orientation (Fig. 45.2E) and of reaching direction (Fig. 45.2F).

These results demonstrate that optic ataxia may be lateralized, occurring only when the patient is required to point to targets in one visual hemifield or only when one hand is used for pointing. In cases where the deficit is restricted to a single hemifield or a single arm, the affected hemifield or arm is usually opposite the injured parietal lobe. This is expected because the parietal lobe of each hemisphere mediates communication between more posterior visual areas (which represent the contralateral half of visual space) and more anterior motor areas (which represent the contralateral arm). There are, however, cases in which the problem is bilateral or occurs for specific combinations of arm and hemifield. Optic ataxia is not simply a problem with visuospatial perception, as indicated by the fact that performance with one arm may be perfectly normal. Nor is it simply a motor problem, as indicated by the fact that patients unable to reach accurately for visual targets can commonly touch points on their own bodies accurately under proprioceptive guidance. Optic ataxia is best characterized as an inability to use visuospatial information to guide arm movements.

Constructional Apraxia: Inability to Reproduce the Spatial Relations of Parts

After injury to the parietal cortex, some patients experience difficulty in apprehending the spatial relations among elements in a scene. They cannot judge whether two objects are of the same size or which of two objects is closer. When asked to copy simple line drawings or to draw from memory, they omit or transpose parts. When attempting to reproduce with an arrangement of matchsticks a visible model in the shape of a four-pointed star, they make gross

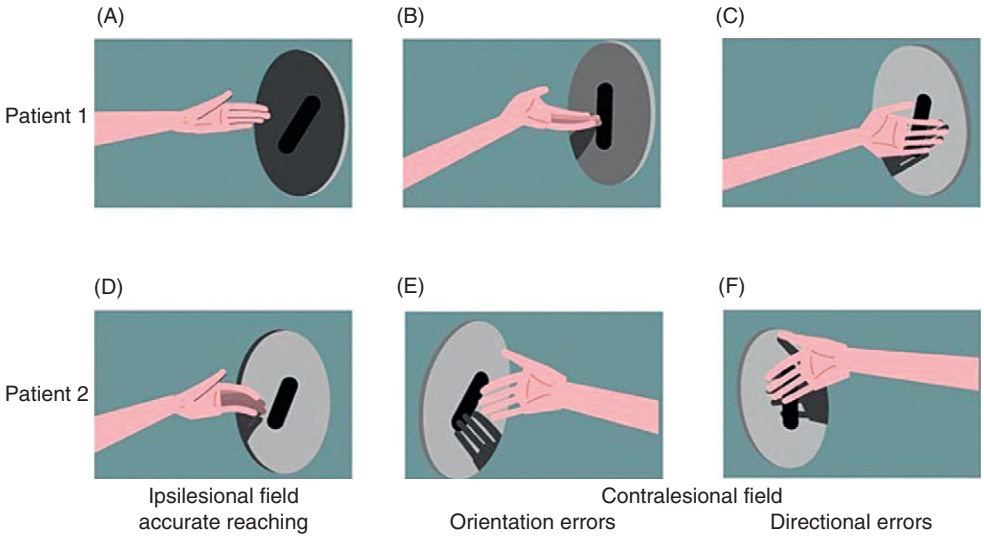


FIGURE 45.2 Optic ataxia. Patients can reach accurately (A and D) when using the ipsilesional hand in the ipsilesional visual field, but they make errors of hand orientation (B and E) and of direction (C and F) when reaching into the contralesional visual field. The top and bottom rows show the performance of two different patients. *From Perenin and Vighetto (1988).*

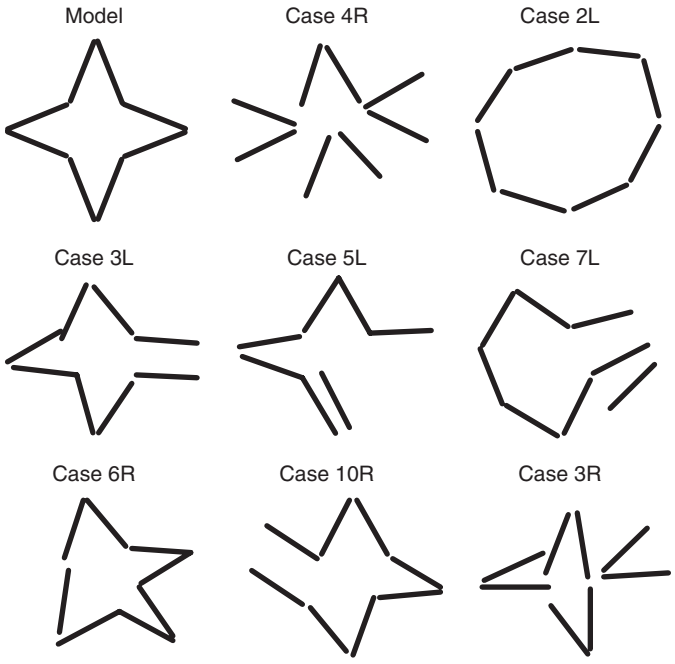


FIGURE 45.3 Constructional apraxia. Patients with parietal lobe injury make gross errors when attempting to arrange matchsticks to resemble a visible model with the shape of a four-pointed star. Cases are labeled R or L to indicate that the site of injury was in the right or left hemisphere. *From McFie and Zangwill (1960).*

errors (Fig. 45.3). The problem does not lie in a loss of fine motor control, as evidenced by the local precision with which they can arrange the matchsticks in globally incorrect configurations. The problem might arise from an inability to organize complex actions in space under visual guidance. Indeed the traditional term "apraxia" implies flawed organization of voluntary actions. It appears, however, that the difficulties arise at the stage of spatial perception rather than at the subsequent stage of motor execution. For example, in one patient with parietal-lobe injury, "difficulties of writing in print were entirely derived from his inability to locate letters and numbers on the page under visual control. Writing cursive with lowercase letters was good, since this does not require specifically visual positioning of each letter" (Vaina, 1994). Constructional apraxia, by this interpretation, is a deficit of visuospatial cognition (which depends on processing where things are relative to each other) and thereby stands in contrast to optic ataxia, a deficit of visuomotor guidance (which depends on processing where things are relative to our own bodies).

Hemispatial Neglect: Unawareness of the Contralesional Half of Space

Hemispatial neglect is a classic symptom of injury to the parietal cortex. Its key manifestation is a lateralized failure of spatial awareness. The most common form of neglect arises from damage to the right parietal lobe and is manifested as a failure to detect objects in the left half of space. Patients with neglect experience problems in daily life, such as colliding with obstacles on the contralesional side of the body or mistakenly identifying letters on the contralesional side of a written word. When asked to copy pictures or to draw simple objects from memory, they leave out details on the affected side. When asked to say whether two objects are the same or different, they tend to indicate that two dissimilar objects are the same if the differentiating details are on the contralesional side.

Neglect involves more than just a defect of attention to contralesional sensory events. For example, patients fail to report detail on the left half of an object even when they are asked to form a mental representation of the object by viewing it one part at a time as it passes slowly behind a vertical slit. In a famous set of experiments, patients were asked to imagine themselves in a familiar public setting and then to report what they would be able to see around them (Bisiach & Luzzatti, 1978). When they imagined standing at one end of the town square, patients described the buildings on the right side and failed to report buildings on the left. However, when instructed to imagine themselves standing at the other end of the square, facing in the opposite direction, they now failed to mention buildings on the left that they had described just moments before and were able to describe buildings on the right that they had previously omitted. These patients were unable to form a mental representation of the contralesional half of space even when the requirement to direct attention to sensory input on that side was altogether removed.

The fact that neglect affects the half of space opposite the injured hemisphere suggests that each parietal lobe must represent the opposite half of space. On the face of it, this proposition does not sound particularly surprising. Most visual areas in each hemisphere represent the opposite half of the visual field, most somatosensory areas represent the opposite half of the skin surface, and most motor areas represent muscles on the opposite half of the body. Injury to these areas leads to a sensory loss or motor impairment that affects the opposite half of a functional space defined with respect to some anatomical reference frame (the retina, the skin, or the muscles). In contrast, injury to the parietal cortex gives rise to a neglect that may be defined with respect to any of several spatial reference frames. These fall into two broad classes: egocentric reference frames, in which locations are represented relative to the observer, and allocentric reference frames, in which locations are represented in coordinates extrinsic to the observer. Examples of egocentric reference frames include those centered on the hand or the body. Allocentric reference frames include those centered on an environmental landmark or an object of interest. Patients with neglect often show deficits with respect to more than one reference frame. The following sections describe evidence for impairments relative to multiple spatial reference frames in neglect.

A patient with left hemispatial neglect, looking straight ahead at the center of some object, will tend not to see detail on its left side. This observation is open to several interpretations. The simplest interpretation is that stimuli presented in the left visual field tend not to be registered. However, the fundamental problem might actually be with registering stimuli that are to the left of the head, the torso or the object itself. These possibilities cannot be distinguished without disentangling the various reference frames. Experiments aimed at identifying the spatial reference frame in neglect indicate that a patient's failure to detect a stimulus is affected not only by its location relative to the retina but also by its location relative to the object or array within which it is contained, by its location relative to the body, and even by its location relative to a gravitationally defined reference frame. Patients span a continuum from exhibiting predominantly egocentric to predominantly allocentric neglect.

In many patients, a component of neglect is object centered. If these patients are presented with an image anywhere in the visual field, they tend to ignore its left side. In one experiment, patients with left hemispatial neglect were required to maintain fixation on a small spot while chimeric faces (images formed by joining at the midline half-images representing the faces of two people) were presented at various visual field locations. Their reports of what they saw were based predominantly on the right halves of the chimeric images. This was true even when the entire composite face was presented within the right visual hemifield. In a second experiment, patients with left hemifield neglect were required to maintain fixation on a small spot while four stimuli in a horizontal row were presented. One of the stimuli was a letter that was to be named aloud. When the letter was at the leftmost location within the array, patients were slowest to name it, even when the entire array was in the right visual field.

A particularly dramatic way of demonstrating the object-centered nature of neglect is to ask patients to make copies of simple line drawings (Marshall & Halligan, 1993). When asked to copy a picture of two flowers, a patient with right parietal lobe damage saw and copied each flower but omitted the petals on the left half of each, thus showing evidence of left object-centered neglect (Fig. 45.4A). When the same two flowers were joined by a common stem, the patient saw and copied the plant, but omitted details on its left side, including all of the leftmost flower, thus showing neglect for the left half of the larger composite object (Fig. 45.4B).

Patients with parietal lobe damage sometimes exhibit neglect defined with respect to allocentric reference frames other than those centered on an object. The role of a gravitational reference frame in neglect has emerged from studies in which patients face a display screen either while sitting upright or while lying on their side (Ladavas, 1987).

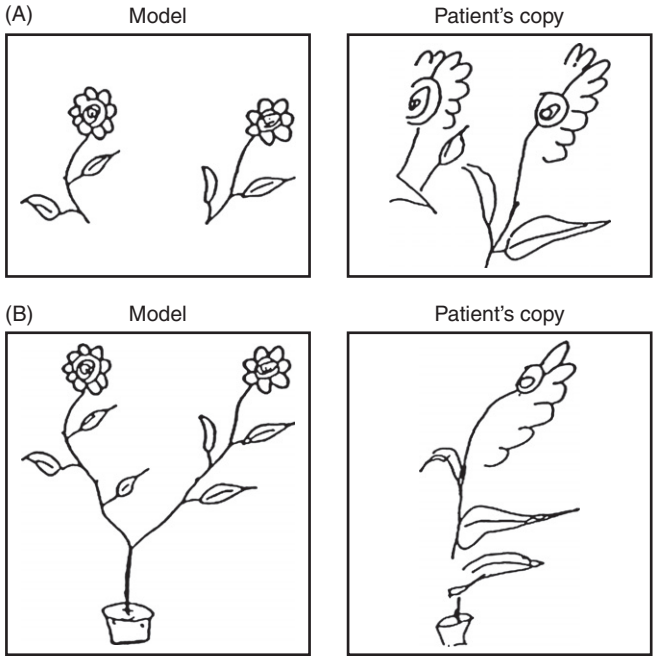


FIGURE 45.4 Object-centered neglect. When asked to copy the two drawings on the left, a patient made the two copies on the right. Detail is omitted from the left half of each object rather than from the left half of the drawing as a whole. *From Marshall and Halligan (1993).*

When the patient sits upright, the right and left halves of the screen coincide with the right and left retinal visual fields. However, when the patient lies on one side while facing the screen, the situation is changed. Reclining on the right side, the patient sees the right and left halves of the screen as being in the upper and lower visual fields, respectively. Applying this procedure to patients with left hemispatial neglect enables the investigator to pose the question: Is the neglect specific for the left retinal visual field or is it specific for the left half of the screen? Neglect turns out to depend in part on each of these factors.

Further evidence that the parietal cortex represents the locations of objects relative to an external framework comes from the observation that neglect, in some patients, is restricted to stimuli within a certain range of distances from the body. The form of neglect termed peripersonal or proximal is specific for stimuli in the immediate vicinity of the body. Another form of neglect, termed extrapersonal, is specific for more distant stimuli. The implication of these findings is that there is a localization of function within the parietal cortex and that neurons in discrete areas process sensory input from objects at different distances. Discrete areas of the parietal cortex are known to be specialized for relaying sensory information to different motor systems. Sensory input from peripersonal space is uniquely significant for guiding reaching movements. In contrast, sensory input from greater distances (extrapersonal space) is useful primarily for the control of eye movements.

Lesions of the Parietal Cortex in Monkeys Produce Spatial Impairments

The original distinction between dorsal and ventral processing streams in the primate visual system was based in part on differences in the effects of lesions of the parietal cortex, the terminus of the dorsal stream, and inferior temporal cortex, the terminus of the ventral stream (Ungerleider & Mishkin, 1982). Monkeys with parietal lesions exhibit a selective impairment of visuospatial cognition. For example, they cannot judge which of two identical objects is closer to a visual landmark. In contrast, lesions of inferior temporal cortex produce deficits in visual pattern recognition (see Chapter 44).

In addition to visuospatial cognitive deficits, monkeys with parietal lesions also exhibit deficits of visuomotor guidance. After parietal lesions, monkeys have difficulty directing eye movements toward targets in the hemispace opposite the side of the lesion (Lynch & McLaren, 1989). Further, when two targets are presented simultaneously in the ipsilesional and contralesional fields, monkeys with parietal lesions tend to ignore the contralesional stimulus, an effect termed visual extinction. Parietal lesions also interfere with the ability of a monkey to reach toward an object. Lesions confined to the inferior parietal cortex impair reaching with the contralesional limb toward a target in contralesional space. Lesions extending across the intraparietal sulcus to include the superior parietal cortex produce impairments in reaching with the contralesional arm into either half of space. In sum, monkeys with parietal lesions exhibit impairments both of visuospatial cognition and of visuomotor guidance.

Parietal Neurons Encode the Spatial Attributes of Objects and Actions

To understand more precisely how the parietal cortex contributes to spatial cognition, several groups of investigators have measured the electrical activity of single neurons during the performance of spatial tasks. These studies were carried out in alert monkeys trained to make movements to visual targets. Because brain tissue itself has no sensory receptors, microelectrodes can be introduced into the brain without disturbing the performance of an animal on a task. By recording neuronal activity during specially designed tasks, neural activity can be related directly to the sensory, cognitive, and motor processes that underlie spatial behavior. The following sections describe how neurons in specific areas within the parietal cortex are selectively activated during spatial tasks and how they contribute to spatial representation and cognition.

Area VIP: Head-Centered Visual and Somatosensory Responses

The ventral intraparietal area (area VIP) is a functionally and anatomically defined area that lies at the junction of the visual and somatosenory districts within parietal cortex. Individual neurons respond strongly to moving stimuli and are selective for both the speed and the direction of the stimulus (Colby, Duhamel, & Goldberg, 1993). In this respect, VIP neurons are similar to those in other dorsal stream visual areas that process stimulus motion, especially areas MT and MST (see Chapter 26). An important finding in area VIP is that most of these visually responsive neurons also respond to somatosensory stimuli (Duhamel, Colby, & Goldberg, 1991). These neurons are truly bimodal in the sense that they can be driven equally well by either a visual or a somatosensory stimulus. Most of these bimodal neurons have somatosensory receptive fields on the head or face that match the location of their visual receptive fields. This correspondence is illustrated in Figure 45.5, which shows visual and somatosensory receptive fields for several individual VIP neurons.

Three kinds of correspondence are found in VIP neurons. First, for each neuron, receptive fields in each modality match in location. For example, a neuron that responds to a visual stimulus in the upper left visual field also responds when the left brow is touched, and a neuron with a foveal visual receptive field also responds to cutaneous stimulation around the mouth. Second, receptive fields match in size. For example, neurons with visual receptive fields around the fovea and somatosensory receptive fields around the mouth have especially small receptive fields in both modalities. This makes sense becase the fovea and the mouth are both regions of high spatial acuity. The third type of correspondence is in the preferred directions of movement. For example, a neuron responsive to a visual stimulus moving toward the right also responds when a small probe is brushed lightly across the face of the monkey moving to the right. In sum, visual and somatosensory receptive fields for individual VIP neurons match in location, size, and directional preference.

The correspondence in receptive field locations across modalities immediately raises a question: What happens to the relative locations of the visual and somatosensory receptive fields when the eyes move? If the visual receptive field were retinotopic, it would move when the eyes do; if the somatosensory receptive field were somatotopic, it would remain in place when the eyes move. Consequently, the correspondence in location between the visual and somatosensory receptive fields would be broken by eye movements.

The answer is that visual receptive fields are linked to the skin surface and not to the retina. A neuron that

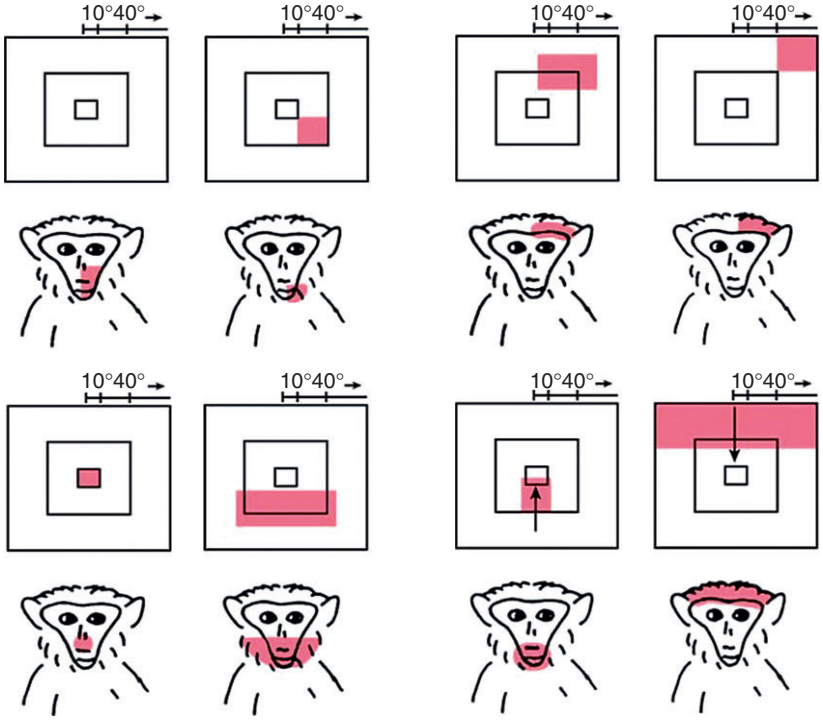


FIGURE 45.5 Matching visual and somatosensory receptive fields of neurons from area VIP. On each outline of the head of the monkey, the patch of color corresponds to the somatosensory receptive field of one neuron. On the square above the head, the patch of texture corresponds to the visual receptive field of the same neuron. The square can be thought of as a screen placed in front of the monkey so that its center was directly ahead of the monkey's eyes at the time of visual testing. *From Duhamel et al. (1991) with permission of Oxford University Press.*

responds best to a visual stimulus approaching the forehead and has a matching somatosensory receptive field on the brow will continue to respond best to a visual stimulus moving toward the brow regardless of where the monkey is looking. Thus, both visual and somatosensory receptive fields are defined with respect to the skin surface. In this sense, the receptive fields are head centered: a given VIP neuron responds to stimulation of a certain portion of the skin surface and to the visual stimulus aligned with it no matter which part of the retina the image falls on.

In sum, neurons in area VIP encode bimodal sensory information in a head-centered representation of space. This kind of representation would be most useful for guiding head movements. Anatomical studies indicate that area VIP sends information to premotor cortex including the specific region that is involved in generating head movements.

Area LIP: Retina-Centered Attentional Signals

The lateral intraparietal area (area LIP) is a functionally defined area within the visual district of parietal cortex. Neurons in LIP exhibit several different kinds of activity during spatial tasks (Colby, Duhamel, & Goldberg, 1996). First, LIP neurons, like neurons in striate and extrastriate visual cortex (see Chapter 26), respond to the onset of a visual stimulus in the receptive field of the neuron. Second, these visual responses are enhanced when the monkey attends to the stimulus: the amplitude of the visual response is increased when the stimulus or stimulus location becomes the focus of attention. This enhancement occurs no matter what kind of movement the animal will use to respond to the stimulus. Regardless of whether the task requires a hand movement or an eye movement or requires that the monkey refrain from moving toward the stimulus, the visual response of an LIP neuron becomes stronger when the stimulus is made behaviorally relevant. This means that the same physical stimulus arriving at the retina can evoke very different responses in cortex as a result of spatial attention.

A third important feature of LIP neuron activity is the sustained responses observed when the monkey must remember the location at which the stimulus appeared. In this task, a stimulus is flashed only briefly, but the neuron continues to fire for several seconds after the stimulus is gone, as though it were holding a memory trace of the target location. A particularly intriguing question in understanding spatial representation concerns the fate of this memory-related activity in area LIP following an eye movement, as will be described later. A fourth kind of activation commonly observed in LIP neurons is related to performance of a saccade—a rapid eye movement—toward the receptive field. Some LIP neurons fire just before the monkey initiates a saccade that will move the fovea onto a target presented in the receptive field. LIP neurons have overlapping sensory and motor fields, just like neurons in the superior colliculus (see Chapter 32).

Finally, LIP neuron activity can be modulated by the position of the eye in the orbit (Andersen, Bracewell, Barash, Gnadt, & Fogassi, 1990). For instance, the visual response of a given cell may become larger when the monkey is looking toward the left part of the screen than when it is looking toward the right. This property is of interest because it suggests that neurons in area LIP contribute to spatial representations that go beyond a simple replication of the retinal map.

In sum, individual LIP neurons have receptive fields at particular retinal locations; they carry visual-, memory-, and saccade-related signals; and their activity can be modulated by attention and eye position. Activity in area LIP cannot be characterized as a simple visual or motor signal. It is probably closer to the truth to think of the firing of an LIP neuron as reflecting how much attention is currently being directed to the location occupied by its receptive field.

Area LIP: Updating of Spatial Information during Eye Movements

Every time we move our eyes, each object in our surroundings activates a new set of retinal neurons. Despite this ever changing input, we experience a stable visual world. How is this possible? More than a century ago, Helmholtz proposed that the reason the world appears to stay still when we move our eyes is that the "effort of will" involved in making a saccade simultaneously adjusts our perception to take that specific eye movement into account. He suggested that when a motor command to move the eyes is issued, a copy of that command, or corollary discharge, is sent to brain areas responsible for generating our internal image of the world. This image is then updated so as to be aligned with the new visual information that will arrive in the cortex after the eye movement. A simple experiment shows that Helmholtz's account must be essentially true. When the retina is displaced by pressing on the eye, the world does seem to move, presumably because there is no corollary discharge. Without that internal knowledge of the intended eye movement, there is no way to update the spatial representation of the world around us.

Neurons in area LIP contribute to this updating of the internal image (Duhamel, Colby, & Goldberg, 1992). The experiment illustrated in Figure 45.6 shows that the memory trace of a stimulus is updated when the eyes move. The activity of a single LIP neuron was recorded under three different conditions. In the first set of trials, the monkey looked at a fixed point on the screen while a stimulus

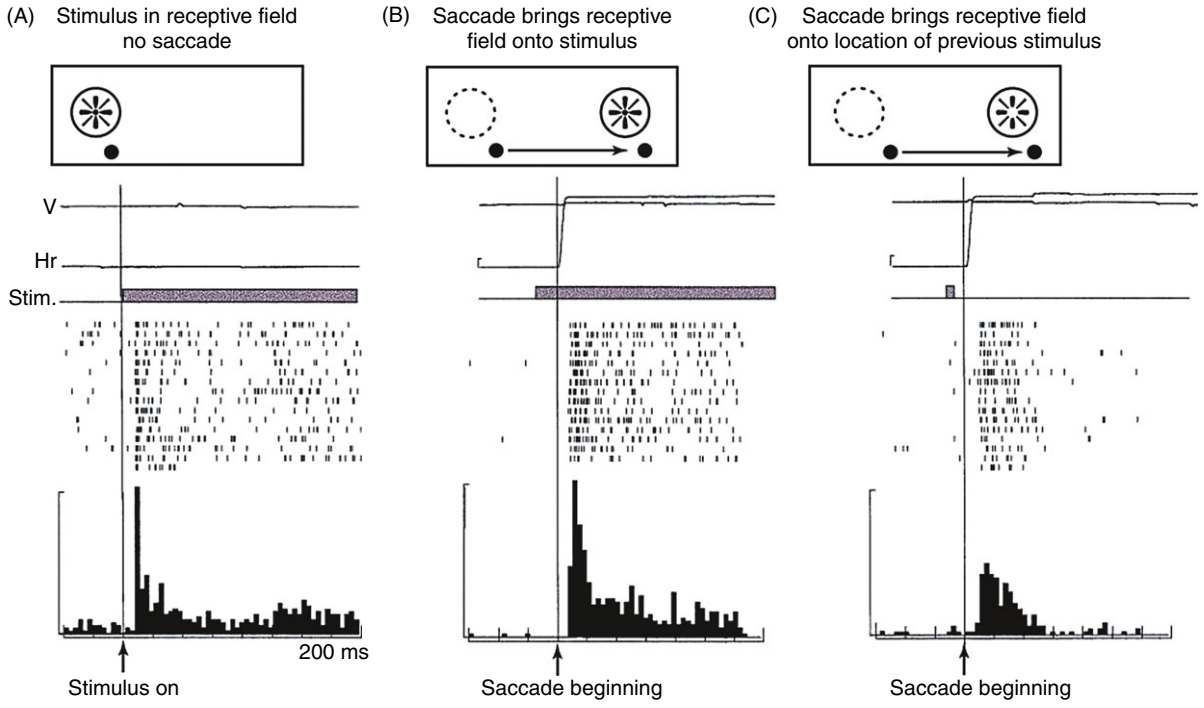


FIGURE 45.6 Remapping of visual memory trace activity in area LIP. The activity of a single neuron was recorded under three different conditions. (A) Simple visual response to a constant stimulus in the receptive field, presented while the monkey is fixating. The rasters and histogram are aligned on the time of stimulus onset. (B) Response following a saccade that brings the receptive field onto the location of a constant visual stimulus. (C) Response following a saccade that brings the receptive field onto the location where a stimulus was presented previously. The stimulus is extinguished before the saccade begins so it is never physically present in the receptive field. The neuron responds to the updated memory trace of the stimulus. V, vertical eye position; H, horizontal eye position. *From Duhamel et al. (1992).*

was presented in the receptive field (Fig. 45.6A). In the diagram at the top of Figure 45.6A, the dot is the fixation point, the dashed circle shows the location of the receptive field, and the asterisk represents the visual stimulus. The time lines just below the diagram indicate that the vertical and horizontal eye positions remained steady throughout the trial, demonstrating that the monkey maintained fixation. The stimulus time line shows that the stimulus appeared 400 ms after the beginning of the trial and continued for the entire trial. The raster display below plots the electrical activity of a single LIP neuron in 16 successive trials. In these rasters, each small dot indicates the time at which an action potential occurred, and each horizontal line of dots represents activity in a single trial. In each trial there was a brief initial burst of action potentials shortly after the stimulus appeared, followed by continuing neural activity at a lower rate. The histogram at the bottom of Figure 45.6A shows the average firing rate as a function of time. The visual response in Figure 45.6A is typical of that observed in neurons in many visual areas: the neuron fired when a stimulus appeared in its receptive field.

In the second set of trials (Fig. 45.6B), a visual response occurred when the monkey made an eye movement that brought a stimulus into the receptive field. At the beginning of the trial, the monkey looked at the fixation point on the left, and the rest of the screen was blank. Then, simultaneously, a new fixation point appeared on the right and a visual stimulus appeared above it. The monkey made a saccade from the old fixation point to the new one, indicated by the arrow. The eye movement was straight to the right so only the horizontal eye position trace shows a change. At the end of this saccade, the receptive field had been moved to the screen location containing the visual stimulus. The rasters and histogram in Figure 45.6B are aligned on the time that the saccade began. In each trial, the neuron began to respond after the receptive field had landed on the stimulus. This result is just what would be expected for neurons in any visual area with retinotopic receptive fields.

The novel finding is shown in Figure 45.6C. In this third set of trials, the monkey made a saccade that would bring a stimulus into the receptive field, just as in the second set of trials. The only difference was in the duration of the stimulus, which lasted for a mere 50 ms instead of staying on for the entire trial. As can be seen on the stimulus time line, the stimulus actually disappeared before the saccade began. This means that it was never physically present in the receptive field. Nevertheless, the neuron fired as though there were a stimulus in its receptive field. This result indicates that LIP neurons respond to the memory trace of a previous stimulus. Moreover, the representation of the memory trace is updated at the time of a saccade. The general idea of how a memory trace can be updated is as follows. At the beginning of the trial, while the monkey is looking at the initial fixation point, the onset of the stimulus activates neurons whose receptive fields encompass the stimulated location. Some of these neurons will have tonic activity and continue to respond after stimulus offset, encoding a memory of the location at which the stimulus occurred. When the monkey moves its eyes toward the new fixation point, a copy of the eye movement command is sent to parietal cortex. This corollary discharge causes the active LIP neurons to transmit their activity to a new set of neurons whose receptive fields will encompass the stimulated screen location after the saccade. By means of this transfer, LIP neurons encode the spatially updated memory trace of a previous stimulus.

The significance of this finding lies in what it illustrates about spatial representation in area LIP. It indicates that the internal image is dynamic rather than static. Tonic, memory-related activity in area LIP not only allows neurons to encode a salient location after the stimulus is gone but also allows for dynamic remapping of visual information in conjunction with eye movements. This updating of the internal visual image has specific consequences for spatial representation in the parietal cortex. Instead of spatial information being encoded in purely retinotopic coordinates, tied to the specific neurons initially activated by the stimulus, the information is encoded in eye-centered coordinates. This is a subtle distinction but a very important one in generating accurate spatial behavior. Visual information encoded in eye-centered coordinates tells the monkey not just where the stimulus was on the retina when it first appeared, but where it would be on the retina following an intervening eye movement. The result is that the monkey always has accurate information with which to program an eye movement toward a real or a remembered target. These phenomena probably occur in the human brain as well, as indicated by the observation that parietal injury can interfere with the ability to keep accurate track of a remembered location during a saccade.

In sum, neurons in area LIP encode the locations of present and remembered visual stimuli in an eye-centered reference frame. This observation suggests that area LIP is involved in eye movement control. It fits with the fact that LIP is strongly and directly connected to other brain centers involved in the control of eye movements, including the frontal eye field and superior colliculus (see Chapter 32).

Area LIP Is Downstream from Spatial Decision Processes

Spatially selective neuronal activity in area LIP is graded in strength. If the monkey's attention to a given location or intention to make a saccade in a given direction increases in strength, then the firing of LIP neurons

with response fields in that direction also increases. This has been demonstrated in decision tasks—tasks requiring the monkey to decide which of several possible saccades to make. A decision can be either sensory-based or value-based. In a sensory decision task, the monkey must judge whether a visual display contains dots moving to the right or left and must report the judgment by executing a rightward or leftward saccade (Roitman & Shadlen, 2002). If the motion stimulus is weak, requiring the monkey to integrate sensory information over time, the neuronal signal reflecting the decision develops slowly. If the motion stimulus is strong, allowing the monkey to make an instantaneous decision, the neuronal signal rises sharply. Thus, the evolution of the monkey's decision is reflected in the activity of spatially selective LIP neurons. In a value-based decision task, the monkey chooses between a saccade to the right and a saccade to the left on the basis of prior experience indicating that one of the saccades is more likely than the other to yield a valued juice reward (Sugrue, Corrado, & Newsome, 2004). During the preparatory period of the trial, LIP neurons fire strongly if the target in their response field is associated with a valuable reward. In this task the activity of spatially selective LIP neurons reflects the value of the target. These two findings do not necessarily indicate that LIP plays a direct role in making sensory or value-based decisions. They can also be explained on the assumption that LIP neurons mediate attention or motor preparation and that they fire more strongly when attention or motor preparation is more intense (Maunsell, 2004).

Summary

The parietal cortex plays a critical role in spatial awareness. Injury to the parietal cortex in humans and monkeys leads to deficits of visuomotor guidance and visuospatial cognition. Parietal neurons in monkeys construct a representation of space by combining signals from multiple sensory and motor modalities. Neurons in different divisions of parietal cortex represent locations relative to several body-centered reference frames including ones centered on the eye and head. Signals carried by these neurons are useful for visuomotor guidance. Recent studies have suggested that parietal neurons also represent locations relative to object-centered reference frames in the service of visuospatial cognition (Box 45.1).

FRONTAL CORTEX

Frontal Cortex Contributes to Voluntary Movement and the Control of Behavior

The frontal lobe is involved in spatial functions as a natural result of its being involved in behavioral control. The three main divisions of the frontal lobe are the primary motor cortex located on the precentral gyrus; the premotor cortex, located in front of the primary motor cortex; and the prefrontal cortex. The primary motor, premotor, and prefrontal areas all contribute to behavioral control, but they differ with respect to the quality of their contribution. This difference is seen in the effects of brain injury. Injury to the primary motor cortex leads to weakness and

BOX 45.1

AREA 7A NEURONS CARRY OBJECT-CENTERED SIGNALS IN A CONSTRUCTIONAL TASK

Many neurons in the monkey parietal cortex represent the locations of visual stimuli relative to parts of the body such as the eye (in area LIP) or head (in area VIP). The body-centered spatial information that they carry is potentially useful for visual guidance of actions such as looking at and reaching for things. Visuospatial cognition, however, requires understanding of where things are relative to each other rather than to the body. It is an interesting question whether monkeys are capable of visuospatial cognition and if so what neural mechanisms underlie it (Caminiti et al., 2010). This issue was clarified by a recent study indicating that monkeys indeed can perform a constructional task requiring the use of object-centered spatial information and that parietal neurons carry object-centered spatial signals during performance of the task (Chafee, Crowe, Averbeck, & Georgopoulos, 2005; Chafee, Averbeck, & Crowe, 2007). On each trial, the monkeys looked at the center of a monitor while two figures appeared in succession. First a simple block figure appeared (model in Fig. 45.7A). Then another block figure formed by removing one block from the model appeared (incomplete copy in Fig. 45.7A). Given a choice of two new blocks (choice array in Fig. 45.7A), the monkeys had to decide which one needed to be slid in horizontally to transform the incomplete copy into a duplicate of the model. On recording from parietal Area 7a, the experimenters found that neurons fired while the monkeys were looking at the incomplete copy and that their firing encoded the location of the missing part. For example, the neuron of Figure 45.7 fired more strongly when the missing part was on the left (B, D, and F) than when it was on the right (C, E, and G). They signaled the location of the missing part relative to the object, not relative to the monkey. The fact that neurons in Area 7a carry object-centered

BOX 45.1 (*cont'd*)

spatial information is consonant with connectional anatomy. Area 7a differs from other parietal areas in being connected strongly to prefrontal cortex, which serves cognitive functions, rather than to premotor cortex, which mediates visuomotor guidance. The loss of neurons with object-centered spatial selectivity might underlie symptoms of parietal injury in humans, including constructional apraxia (Fig. 45.3) and object-centered neglect (Fig. 45.4).

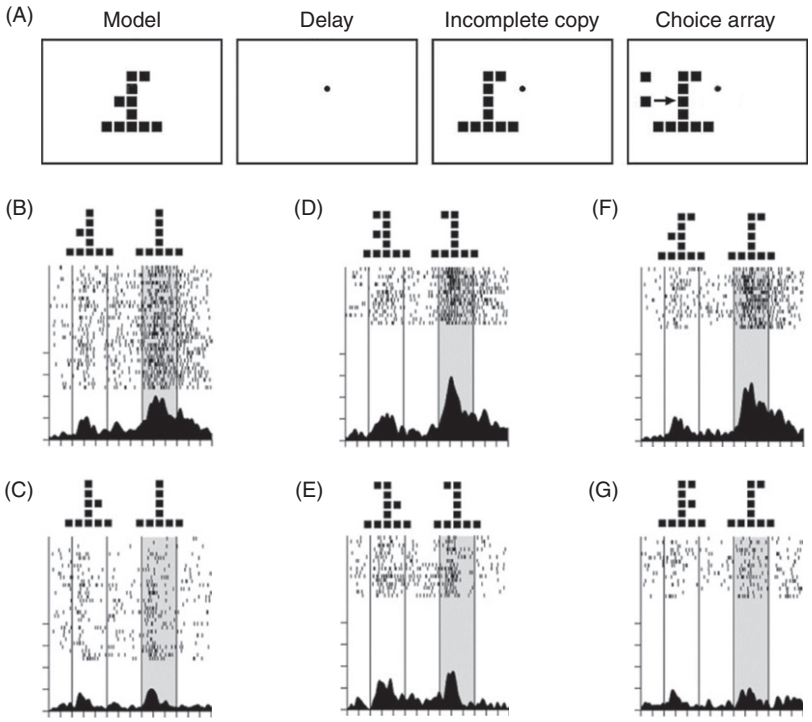


FIGURE 45.7 Data from a neuron in parietal Area 7a with object-centered spatial selectivity. (A) While the monkey looked at a fixation point at the center of the screen (the central dot in each panel), a model block figure appeared followed, after a delay, by an incomplete copy from which one block had been removed. In this example, the block protruding halfway up the model on the left side was missing. A choice array consisting of two free blocks then appeared next to the incomplete copy. The monkey had to choose which free block to slide in horizontally so as to transform the incomplete copy into the original model (arrow). In this example, the lower block was the correct choice. (B–G) Data from an Area 7a neuron. This neuron fired during the period when the monkey was viewing the incomplete model. It fired much more strongly if the missing block was on the left of the model than if it was on the right. Note that the conditions consitute matched pairs in which monkey was viewing an identical incomplete copy (B–C, D–E, and F–G). The only thing that differed between the two conditions in a pair was the monkey's awareness of the object-centered location (left or right) of the missing part. *From Chafee et al. (2005).*

paralysis of the contralateral muscles. Injury to the premotor cortex leads to difficulty in producing movements under certain circumstances—for example, when the patient is asked to mime the use of a tool or to learn arbitrary associations between stimuli and responses. Finally, injury to the prefrontal cortex results in a syndrome characterized by reduced drive, reduced ability to form and carry out plans, and an impairment of working memory (see Chapter 50). These effects indicate that progressively more anterior parts of the frontal lobe contribute to progressively more abstract aspects of behavioral control. We describe following findings from the monkey that cast light on how space is represented in primary motor, premotor, and prefrontal cortex.

Neurons of Primary Motor Cortex Represent Movement Direction Relative to Both Intrinsic and Extrinsic Frames

The primary motor cortex contains a map of the muscles of the body in which the leg is represented medially, the head laterally, and other body parts at intermediate locations. Within this map are patches of neurons that represent different muscles. Neurons within a given patch receive proprioceptive input from a muscle or small group of synergistic muscles and send their output back to that muscle or group of synergists by way of a multisynaptic pathway through the brainstem and spinal cord. There have been many studies in which the electrical activity of

neurons in the primary motor cortex is monitored while animals move (see Chapter 29). The general conclusion of these studies is that neurons in the primary motor cortex are active when the corresponding muscles are undergoing active contraction. Neurons in the primary motor cortex probably do more, however, than simply encode the levels of activation of individual muscles. One proposal is that they encode movement trajectories. Every voluntary movement can be described in two quite different but perfectly complementary ways: in terms of the lengths of the muscles or in terms of the position of the part of the body being moved. For example, during an arm movement, changes take place both in the lengths of muscles acting on the arm and in the position of the hand. Could it be that neurons in the primary motor cortex encode an extrinsic variable, such as the direction of movement of the hand, rather than an intrinsic variable such as which muscles are contracting?

Evidence for the idea that some neurons of the primary motor cortex encode movement direction relative to an extrinsic frame has come from studies in which monkeys make reaching movements in various directions. Individual neurons are selective for a specific direction of movement: a given neuron may fire most strongly during movements up and to the right and progressively less strongly for movements that deviate from the preferred direction (Schwartz, Kettner, & Georgopoulos, 1988). The patterns of selectivity are well defined, and the preferred directions of different neurons cover the range of possible movements fairly evenly. By recording the activity of the entire population of neurons one could, in principle, quite accurately describe the movement. The question remains whether these neurons are selective for the direction of movement in space or for the specific patterns of muscle activation associated with particular movements. In an elegant study, Kakei, Hoffman, and Strick (1999) recorded from neurons in the primary motor cortex while the monkey moved its hand in a single direction using different combinations of muscles. They discovered that the primary motor cortex contains both neurons selective for movement direction and neurons selective for patterns of muscle activation.

Premotor Neurons Resemble Parietal Neurons in their Patterns of Spatial Selectivity

One of the general functions of premotor cortex is to act as a conduit through which signals generated in parietal cortex control motor output. Connections between parietal and premotor cortex are organized topographically with the result that each functionally distinct subregion of parietal cortex projects particularly strongly to a related region in premotor cortex (Rizzolatti & Lupino, 2001). Premotor neurons are thought to have a more direct involvement in voluntary motor control than parietal neurons. However, they appear to preserve parietal codes for representing the locations of targets and the directions of movements. Thus, the functional properties of neurons in each region of premotor cortex mirror the functional properties of neurons in the related parietal region. For example, neurons in the division of ventral premotor cortex (PMv) linked to parietal area VIP respond both to visual and to somatosensory stimuli and possess head-centered receptive fields (Fogassi et al., 1992). Likewise, neurons in the frontal eye field (FEF), which is linked to Area LIP, possess retina-centered visual receptive fields and exhibit remapping in conjunction with eye movements (Umeno & Goldberg, 1997).

Neurons in the Supplementary Eye Field Encode the Object-Centered Direction of Eye Movements

A particularly interesting form of allocentric spatial representation analogous to that observed in area 7a (Box 45.1) has been observed in the supplementary eye field (SEF). The SEF is a division of the premotor cortex with attentional and oculomotor functions. Neurons in this area fire before and during the execution of saccadic eye movements. Several characteristics of the SEF set it apart from subcortical oculomotor centers and suggest that its contribution to eye movement control occurs at a comparatively abstract level. For example, neurons here fire while the monkey is waiting to make an eye movement in the preferred direction, as well as during the eye movement itself. Moreover, some SEF neurons become especially active when the monkey is learning to associate arbitrary visual cues with particular directions of eye movements.

In monkeys trained to make eye movements to particular locations on an object, SEF neurons encode the direction of the impending eye movement as defined relative to an object-centered reference frame (Olson & Gettner, 1995). Regardless of where an object appears relative to the body, these neurons respond when the animal is planning to make an eye movement to a specific location on that object. An experiment demonstrating this point is shown in Figure 45.8. This figure shows the activity of a single SEF neuron while the monkey performed an eye movement task. At the beginning of each trial, the monkey fixated a dot at the center of a screen. While the monkey fixated, a sample and cue were presented in the right visual field. The cue flashed on either the right or the left end of the short horizontal sample bar. Following extinction of the sample–cue display, the monkey maintained fixation for a brief time. Then, simultaneously, the central spot was extinguished and a target bar appeared at an unpredictable location in the upper visual field of the monkey. The monkey had to make an eye movement to the end of the target bar corresponding to the cued end

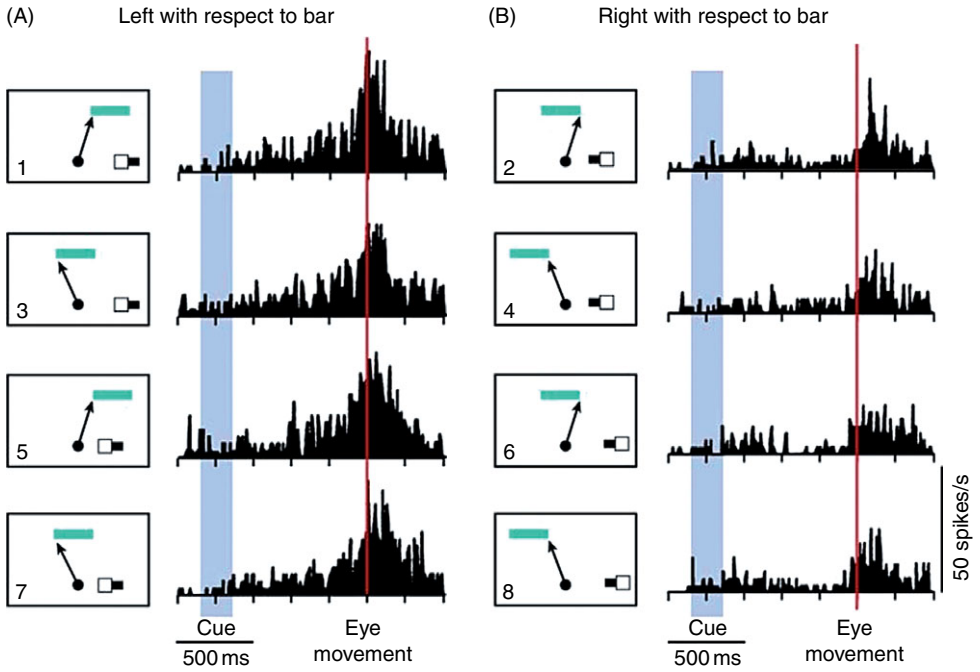


FIGURE 45.8 Data from a neuron in the SEF selective for the object-centered direction of eye movements. The monkey was trained to make eye movements to the right or left end of a horizontal bar. A cue appeared on the small sample bar shown in the lower right of each panel (1–8) to tell the monkey which end of the target bar (top in each panel) was relevant on a given trial. The arrow in the panel next to each histogram indicates the direction of the eye movement. The neuron fired strongly when the eye movement was directed to the left end of the target bar (left column) regardless of whether the physical movement of the eye was up and to the right (panels 1 and 5) or up and to the left (panels 3 and 7). *From Olson and Gettner (1995).*

of the sample bar. Across the set of eight conditions, the object-centered direction of the eye movement was completely independent of its physical direction.

This situation made it possible to ask whether the activity of the neuron was related to the object-centered direction of the movement or to its physical direction. To the right of each panel in Figure 45.8 is shown the average firing rate of a single neuron as a function of time during the trial. Regardless of the direction of the physical movement of the eye, firing was clearly stronger on trials in which the monkey made an eye movement to the left end of the target bar than on trials in which the target was the right end of the bar. For example, in panels 1 and 2, the physical direction of the eye movement was exactly the same, and yet firing was much stronger when the left end of the bar was the target (panel 1) than when the right end of the bar was the target (panel 2). This neuron exhibited object-centered direction selectivity in the sense that it fired most strongly before and during movements to a certain location on an object.

Object-centered direction selectivity serves an important function in natural settings. In scanning the environment, we sometimes look toward locations where things are expected to appear, but which, at the time, contain no detail, for example, the center of a blank screen or the center of an empty doorway. Our eyes are guided to these featureless locations by surrounding features that define them indirectly. It is specifically in these cases that the SEF may contribute to the selection of the target for an eye movement.

The Prefrontal Cortex Mediates Working Memory for Spatial Information

Working memory is required to hold a plan in mind and carry it out step by step (see Chapter 50). The fact that this ability is severely impaired in some patients following prefrontal injury indicates that prefrontal cortex plays a crucial role in working memory. Insofar as plans and working memory have a spatial component, operations carried out by the prefrontal cortex are also spatial in nature. Single neuron-recording experiments in monkeys performing delayed-response tasks have demonstrated the importance of the dorsolateral prefrontal cortex for both spatial and nonspatial forms of working memory (Funahashi, Bruce, & Goldman-Rakic, 1991). A delayed-response task consists of several trials, each several seconds long. At the beginning of each trial, a cue is presented briefly, instructing the monkey which response to make, but the reponse must be withheld until the end of the trial. Delayed-response tasks can be designed so that both the cues (for example, spots flashed to the right or left of fixation) and the responses (for example, eye movements to the right or left) are spatial. In the context of these spatial

tasks, prefrontal neurons are active during the period between the cue and the signal to respond, when the monkey is holding spatial information in working memory. The simple interpretation of this pattern of activity is that it is a neural correlate of the monkey actively remembering the cue and holding in mind the intended response. This interpretation is consistent with results from brain imaging studies demonstrating prefrontal activation in humans performing working memory tasks.

The idea that the prefrontal cortex mediates spatial working memory has received further support from lesion experiments in monkeys. After injury to or inactivation of specific locations in the prefrontal cortex, monkeys are impaired at remembering locations in the opposite half of space. Their delayed responses are spatially inaccurate and the inaccuracy is exacerbated by making delays longer. An experiment demonstrating the importance of the prefrontal cortex for spatial working memory is illustrated in Figure 45.9, which shows behavioral data from a monkey trained to perform an oculomotor delayed-response task (Sawaguchi & Goldman-Rakic, 1994). At the beginning of each trial, the monkey fixated a spot at the center of the screen. While the monkey maintained fixation, a cue was flashed at one of six possible locations.

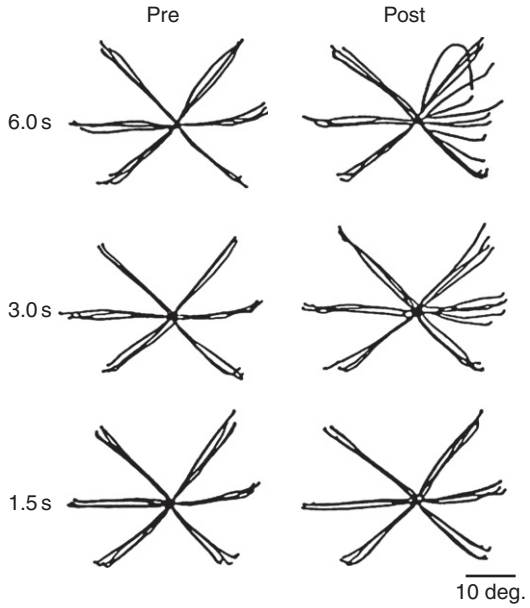


FIGURE 45.9 Inactivation of the left prefrontal cortex disrupts spatial working memory in the right half of space. A monkey was trained to fixate a central spot while a peripheral cue was presented briefly at one of six locations. After an interval of 6, 3, or 1.5 seconds, the fixation spot was extinguished and the monkey made an eye movement to the remembered location. Each set of diverging rays shows the trajectories of eye movements executed under a certain set of conditions. The left column shows eye movement performance before inactivation by a dopamine antagonist. The right column shows performance after inactivation. When the cue was in the right hemifield and the monkey was required to remember its location over long delays, the movements became highly inaccurate. *From Sawaguchi and Goldman-Rakic (1994).*

Following presentation of the cue, a delay of 1.5 to 6 seconds ensued before the monkey was permitted to make an eye movement to the cued location. Data from a normal monkey are shown in the left column of Figure 45.9. In each panel, the six bundles of radiating rays represent the eye trajectories on trials when the cue was at the six different locations. Even when required to remember the cue for 6 seconds, the monkey made accurate eye movements. Data in the right column are from the same monkey after a dopamine antagonist was injected into the left prefrontal cortex. The ability of the monkey to remember the location of the cue remained intact when the cue was in the left (ipsilesional) visual field, as shown by the accurate eye movements. However, performance deteriorated on trials when the cue was in the right (contralesional) visual field. Especially after long delays (top panel), the direction of the eye movement began to deviate from the location of the cue, as if the working memory of the monkey were fading. The fact that this deficit was specific to long delays is noteworthy because it rules out any simple explanation based on interference with visual or motor processes as opposed to working memory itself.

Neural signals related to working memory are observed in object memory tasks as well as spatial tasks (see Chapter 50). In a task in which the monkey had to remember the identity of an object as well as its location, some neurons were selectively activated by remembering the identity of the object independently of its location (Rao, Rainer, & Miller, 1997). Thus, the functions of prefrontal cortex include but are not limited to spatial cognition.

Value Representation and Motivational Modulation in Frontal Cortex

In many areas of the frontal lobe, neurons fire more strongly when a monkey anticipates receiving a larger reward for making a saccade (Roesch & Olson, 2003). The firing of these neurons could signify either of two things. First, it might represent, in an economically meaningful sense, the value of the reward that the monkey expects to receive. The monkey might commit more or less firmly to making the saccade on the basis of how strongly the neurons are firing. Second, it might simply be related to motor planning, attention, or arousal—internal states enhanced when the monkey is more motivated. That these interpretations are different can be illustrated by a simple example. The neck muscles of some monkeys grow more tense when they are planning a saccade in expectation of getting a larger reward (Roesch & Olson, 2004). The tension of the neck muscles obviously reflects motivation and does not represent value in an economically meaningful sense. Whether the activity of frontal neurons represents value or reflects motivation cannot be determined by experiments manipulating solely the size of the expected reward because, as value increases, so does motivation.

This issue has been resolved by manipulating independently the size of the reward promised in the event of the monkey's successfully completing the trial (one or three drops of juice) and the size of the penalty threatened in the event of failure (a one-second or eight-second time-out). A large promised reward and a large threatened penalty have opposite value but both act to increase motivation. Spatially selective neurons in premotor cortex fire more strongly when either a large reward has been promised or a large penalty has been threatened (Fig.

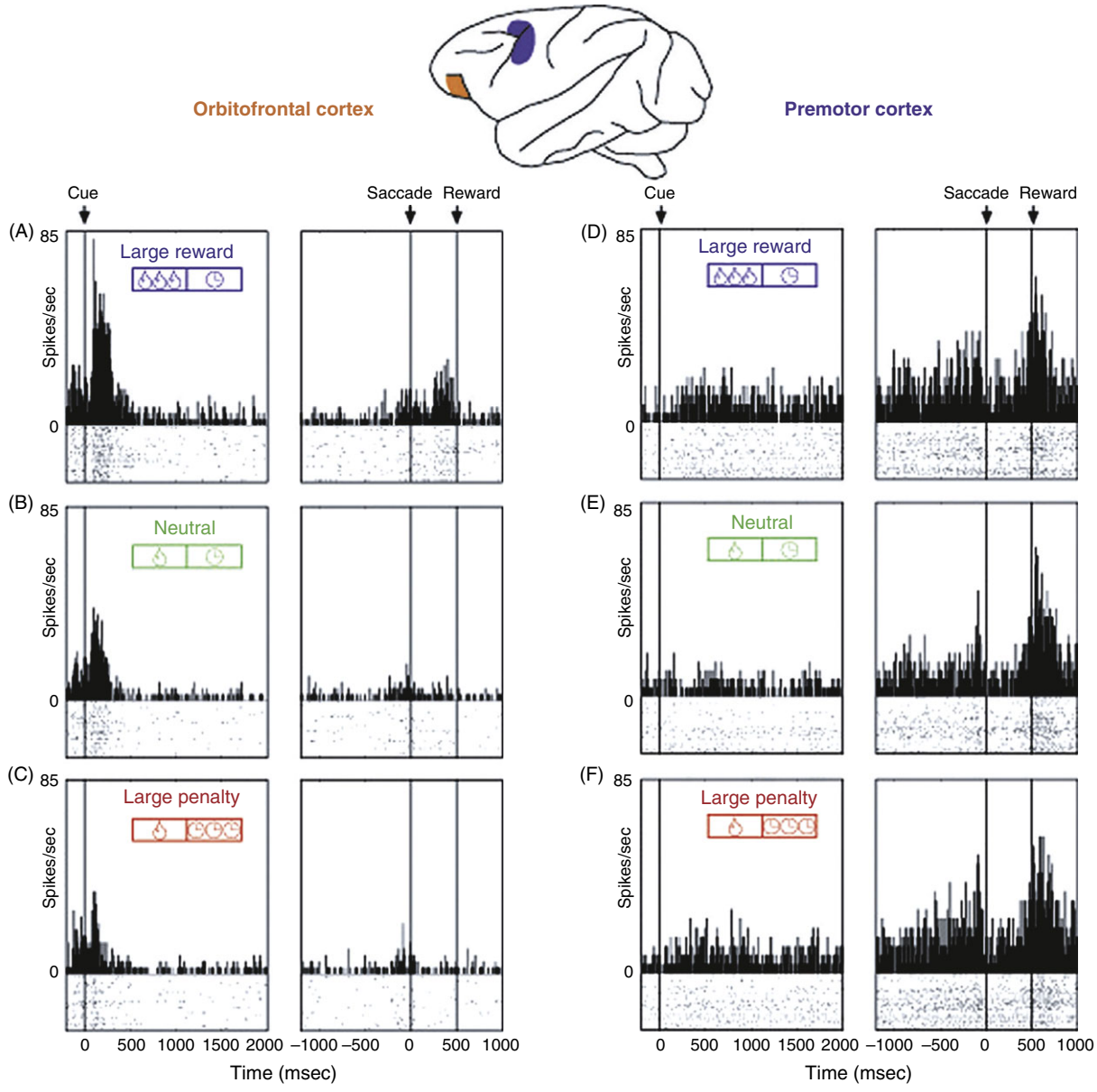


FIGURE 45.10 Evidence that orbitofrontal neurons encode expected value whereas premotor neurons are subject to motivational modulation. A cue presented early in each trial instructed the monkey whether to make a rightward or leftward saccade at the trial's end. It also indicated whether making the instructed saccade would lead to a large or small reward (three drops or one drop of juice). Finally, it indicated whether failure to make the saccade would lead to a large or small penalty (a 1-second or 8-second time-out). After a delay of several seconds, the monkey was permitted to respond. For a saccade in the instructed direction, the monkey received the promised reward. For any other response, the monkey incurred the threatened penalty. The orbitofrontal neuron encoded the value of the outcomes predicted by the cue. It responded most strongly when the combined predicted outcome was best (A: large reward promised and small penalty threatened), at an intermediate level when it was middling (B: small reward promised and small penalty threatened) and least strongly when it was worst (C: small reward promised and large penalty threatened). The premotor neuron fired during the delay period at a rate that reflected how motivated the monkey was. It fired most strongly when the monkey had something significant to work for, either earning a large reward (D) or avoiding a large penalty (F) and least strongly when little was at stake (E). In each panel, time is on the horizontal axis. The histogram represents the average firing rate of the neuron as a function of time during the trial. In the underlying raster display, each horizontal line corresponds to a single trial and each dot represents an action potential. *From Roesch and Olson (2004).*

45.10). In other words, activity in these neurons is modulated by the animal's level of motivation. In contrast, neurons in orbitofrontal cortex fire most strongly when a large reward has been promised and least strongly when a large penalty has been threatened. In other words, they represent anticipated value. The orbitofrontal cortex is a part of the limbic system known to be involved in emotional processes. Presumably, when the monkey considers an action that would lead to an outcome, neurons in orbitofrontal cortex signal the emotion that would be elicited by the outcome, thus representing its value. Value representations in orbitofrontal cortex may give rise to motivational modulation in premotor cortex.

Summary

Neurons in the frontal cortex represent spatial information in conjunction with the role they play in controlling spatially organized behavior. Neurons in the primary motor cortex encode the directions of movements. Neurons in the premotor cortex encode locations relative to numerous body-centered reference frames. SEF neurons encode locations in an allocentric, object-centered representation. Neurons in the prefrontal cortex encode the locations of and identity of objects being held in short-term memory. The motor and cognitive spatial signals carried by neurons in the frontal cortex are subject to motivational modulation.

MEDIAL TEMPORAL LOBE

The Hippocampal System Mediates Declarative Spatial Memory

The hippocampus is an area of primitive cortex, or allocortex, hidden within the medial temporal lobe. It is connected to a set of immediately adjacent temporal-lobe areas, including the perirhinal, entorhinal, and parahippocampal cortices (Zola-Morgan & Squire, 1993). The hippocampal system plays a critical role in the formation of declarative memories in humans (see Chapter 48). Consequently, patients with damage to the hippocampal system sometimes exhibit quite dramatic impairments of spatial memory as manifest in navigational behavior. The noted patient H.M., for example, experienced profound difficulty in learning to navigate through a new neighborhood. Likewise, patient E.P., profoundly amnesic after extensive medial temporal damage resulting from a viral encephalitis, could not learn the layout of a new neighborhood (Teng & Squire, 1999). The deficit in these cases concerned the formation of declarative memories in general rather than some specifically spatial function. Navigation, in itself, was intact, as evidenced by good performance in neighborhoods familiar from before the injury. Moreover, the amnesia extended to non-spatial material, as evidenced by behavior such as failing to recognize examiners to whom the patients had been introduced only minutes before.

Parahippocampal and Retrosplenial Cortex Contribute to Navigation

The retrosplenial and parahippocampal areas are located in the medial temporal lobe close to the hippocampus. They are linked by multisynaptic pathways both to parietal cortex and to the hippocampus. Thus, they are strategically positioned to mediate navigational ability, which depends on both spatial information processing and memory. Studies in humans, based on functional imaging and on the analysis of the behavioral effects of lesions, indicate that these areas may indeed play a special role in environmental navigation (Epstein, 2008). The parahippocampal place area (PPA) and the retrosplenial cortex (RSC) respond with stronger functional activation to images of scenes than to images of faces or objects (see Chapter 44). Moreover, damage to the region of the medial temporal lobe within which they reside leads to topographic disorientation (Barash, 1998). In one form of topographic disorientation, thought to arise from damage to the PPA, individuals become lost because they cannot recognize places that should be familiar (Epstein, 2008). In another form of the disorder, thought to arise from damage to the RSC, individuals are able to recognize familiar places but are unable to orient from the familiar place to a distant destination (Epstein, 2008).

Neurons in the Rat Hippocampal System Encode Place and Heading

Recordings from hippocampal neurons in rats running mazes or exploring open areas have revealed a remarkable degree of spatial selectivity. Many neurons in the hippocampus have place fields: a given neuron will fire most strongly when the rat is within a certain sector of the environment. Different neurons have distinct place fields so that, collectively, they cover the environment (Fig. 45.11B). Place fields are defined relative to prominent environmental landmarks. When a radially symmetric eight-arm maze is rotated relative to a surrounding room containing salient landmarks the place fields remain fixed with respect to the room. Even in darkness, neurons continue to fire when the rat is in their place field, indicating that these responses are not simply visual. If the rat is placed in a new environment, hippocampal neurons develop new place fields rapidly. The location of the place field of a neuron in the new environment is generally unrelated to its location in the old environment. Thus, the hippocampus develops an entirely new place-field map for

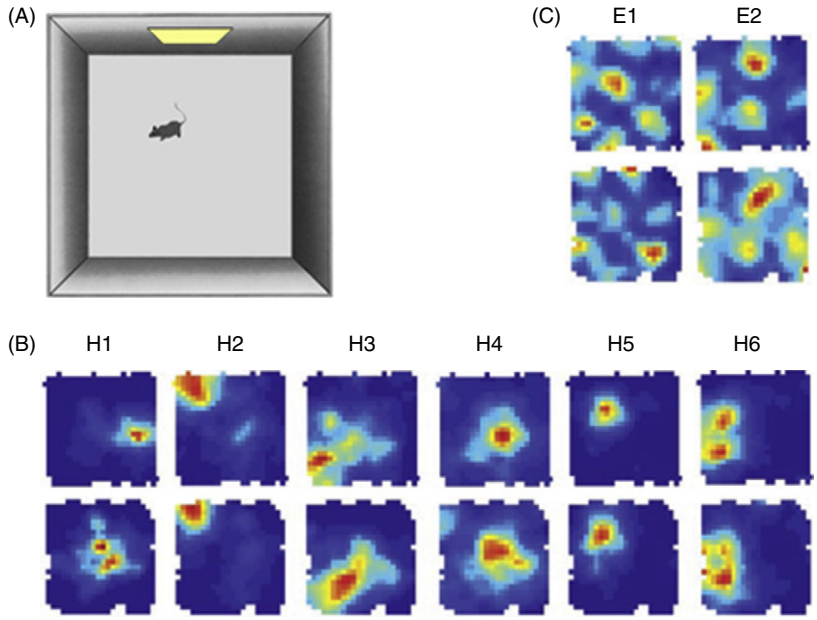


FIGURE 45.11 Neurons in the hippocampal system fire differentially as a function of a rat's location in the environment. (A) In this experiment, the rat moved freely through a square enclosure with low walls in which a card placed on one wall formed a visible landmark. (B) Data from six hippocampal neurons (H1 through H6) with place fields. Each heat map depicts the rate of firing of the neuron as a function of the rat's location in the enclosure (red indicates strong activity and dark blue indicates weak activity). The top row and the bottom row, representing data from separate sessions, are included to indicate the reproducibility of environmental spatial selectivity. (C) Data from two entorhinal grid cells (E1 and E2). Conventions as in (B). Note that each grid cell has multiple sub-fields within the enclosure. *From Fyhn, Molden, Witter, Moser, and Moser (2004).*

each new environment. This is true even in cases where the workspace is changed without any change in the surrounding room—for example, through replacement of an eight-arm radial maze by an open field.

Neurons in other regions of the hippocampal system are sensitive to the direction in which the head is pointing. Sensitivity to heading is common in the postsubiculum, a cortical area adjacent to and closely linked to the hippocampus. Each postsubicular neuron fires most strongly when the head of the rat is pointing in its preferred direction and fires progressively less strongly as the orientation of the head of the rat deviates farther from that direction. In a room that is suddenly darkened, postsubicular neurons remain sensitive to the heading of the rat so long as the rat retains a sense of spatial orientation. When the sense of spatial orientation drifts away from veridicality during prolonged darkness, as evidenced by the occurrence of systematic errors on spatial tests, then the preferred headings of postsubicular neurons exhibit a commensurate drift. These observations establish that neurons of the hippocampal system are sensitive to the rat's perceived heading.

Within limits, rats are able to maintain an accurate sense of place while they move through a dark environment. We know that they can do so both because they are able to find familiar locations and because the firing of their place cells is appropriately updated as they move. The updating of place cell activity depends on an underlying computation often termed path integration. In path integration, the rat's old location and the rat's own movement are added together to give the rat's new location. The computation is easy to describe but the neural machinery mediating it has not been understood. Recent evidence from single-neuron recording studies of the entorhinal cortex has suggested a solution to this problem. The entorhinal cortex, a source of strong and direct input to the hippocampus, contains grid cells (Hafting, Fyhn, Molden, Moser, & Moser, 2005; Moser, Kropff, & Moser, 2008). In contrast to a hippocampal place cell, which possesses a single place field in any given environment, an entorhinal grid cell possesses a multitude of firing fields covering the environment in a strikingly regular hexagonal grid (Fig. 45.11C). The firing fields of different grid cells cover the environment at a variety of scales and phases so that for each location in the environment there is a unique pattern of activity across the population of grid cells. Thus, grid cells in the entorhinal cortex encode information about the rat's current location and could relay this information to place cells in the hippocampus. The question remains: why should place information be represented in such a complex code at the level of entorhinal cortex? Soon after the discovery of grid cells, computational neuroscientists realized that representing environmental locations in a grid code could facilitate the computations that underlie path integration. Path integration has now been implemented in several models of the hippocampal system that incorporate grid cells (Fuhs & Touretzky, 2006; McNaughton, Bataglia, Jensen, Moser, & Moser, 2006).

Summary

The hippocampal system mediates the formation of declarative memories, including memories with spatial content. Injury to the hippocampal system induces a profound anterograde amnesia manifest in the spatial domain as an amnesia for neighborhood layout. In addition to mediating domain-general memory processes, the medial

temporal lobe may make specific contributions to navigation in the environment. In humans, the parahippocampal and retrosplenial cortices contribute to recognizing familiar places and orienting from them to distant goals. In the rat hippocampal system, the existence of place cells, head-direction cells, and grid cells argues for a specific and dedicated role in maintaining and updating a representation of the animal's location and heading in the environment.

SPATIAL COGNITION AND SPATIAL ACTION

The preceding sections described how spatial information is processed in cortical areas that serve distinct functions, such as motor control, attention, working memory, declarative memory, and navigation. Even within the motor system, there appears to be a fractionation of spatial functions in that neurons controlling movements of the eyes, head, and arm represent the locations of visible targets relative to eye-, head-, and hand-centered reference frames, respectively. In addition to these distinctions, there may be another fundamental distinction within brain systems mediating spatial functions. Areas mediating conscious awareness of spatial information may be partially separate from those mediating the spatial guidance of motor behavior. These functions might seem inseparable insofar as one must be aware of the location of a thing in order to look at it or reach for it. However, this is not necessarily the case. An indication that spatial awareness and spatially programmed behavior are distinguishable has come from studies of patients with "blindsight" (Weiskrantz, 1996). This condition arises as a result of injury to the primary visual cortex. Patients experience a scotoma, an area of blindness, in the part of the visual field represented by the injured cortex. The blindness is total in the sense that patients do not report seeing visual stimuli when stimuli are presented within the confines of the scotoma. Nevertheless, if patients are asked to look toward or to reach for an unseen stimulus, choosing the target by guesswork, their responses are directed to the correct location.

Similar findings have been reported in patients with diffuse pathology affecting widespread areas, including the prestriate visual cortex. One such patient, D.F., has been extensively studied in spatial tasks (Goodale et al., 1994). The results reveal an important distinction between spatial performance and spatial cognition. When asked to express spatial judgments (for example, to indicate the size of a visible object by spreading the thumb and forefinger), D.F. performs poorly. Nevertheless, when asked to make visually guided movements (for example, to reach for an object), she accurately adjusts her grip size and hand orientation under visual control to grasp the object efficiently. The fact that intact visuomotor performance coexists with profoundly impaired visuospatial perception in some patients seems to argue for the existence of distinct brain systems specialized for conscious spatial awareness and motor guidance. Many lines of evidence indicate that conscious spatial awareness depends on relatively lateral regions of parietal cortex while motor guidance depends on more medial regions of parietal cortex. Conscious spatial vision is distinct from conscious visual recognition, which depends on occipitotemporal cortex (see Chapter 44).

Summary

Spatial cognition is a function of many brain areas. No one area is uniquely responsible for the ability to carry out spatial tasks. Nevertheless, some generalizations can be made about the part of the problem that is solved in each brain region. The parietal lobe plays a crucial role in many aspects of spatial awareness. It contains numerous areas within each of which neurons carry spatial information relevant to a particular form of behavior.

Neural pathways emanating from parietal cortex relay spatial information to at least three cortical systems that elaborate it for specific purposes. The pathway to premotor cortex is critical for spatial guidance of actions in the workspace surrounding the body. The pathway to prefrontal cortex is critical for cognitive functions including spatial working memory. The pathway to the medial temporal lobe is critical for spatial declarative memory and for navigational ability.

Beneath the seeming unity of spatial experience lies a diversity of specific neural representations. The brain constructs not one but many internal representations of space. A challenge for the future is to understand how these many representations are bound up into that single seamless awareness of space with which each of us is familiar.

References

Andersen, R. A., Bracewell, R. M., Barash, S., Gnadt, J. W., & Fogassi, L. (1990). Eye position effects on visual, memory, and saccade-related activity in areas LIP and 7a of macaque. *Journal of Neuroscience, 10,* 1176–1196.

Barash, J. (1998). A historical review of topographical disorientation and its neuroanatomical correlates. *Journal of Clinical and Experimental Neuropsychology, 20,* 807–827.

Bisiach, E., & Luzzatti, C. (1978). Unilateral neglect of representational space. *Cortex, 14,* 129–133.

Caminiti, R., Chafee, M. V., Bataglia-Mayer, A., Averbeck, B. B., Crowe, D. A., & Georgopoulos, A. P. (2010). Understanding the parietal lobe syndrome from a neuropychological and evolutionary perspective. *European Journal of Neuroscience, 31,* 2320–2340.

Chafee, M. V., Averbeck, B. B., & Crowe, D. A. (2007). Representing spatial relationships in posterior parietal cortex: Single neurons code object-referenced position. *Cerebral Cortex, 17,* 2914–2932.

Chafee, M. V., Crowe, D. A., Averbeck, B. B., & Georgopoulos, A. P. (2005). Neural correlates of spatial judgment during construction in parietal cortex. *Cerebral Cortex, 15,* 1393–1413.

Colby, C. L., & Duhamel, J. -R. (1991). Heterogeneity of extra-striate visual areas and multiple parietal areas in the macaque monkey. *Neuropsychologia, 29,* 517–537.

Colby, C. L., Duhamel, J. -R., & Goldberg, M. E. (1993). Ventral intraparietal area of the macaque: Anatomic location and visual response properties. *Journal of Neurophysiology, 69,* 902–914.

Colby, C. L., Duhamel, J. -R., & Goldberg, M. E. (1996). Visual, pre-saccadic and cognitive activation of single neurons in monkey lateral intraparietal area. *Journal of Neurophysiology, 76,* 2841–2852.

Duhamel, J. -R., Colby, C. L., & Goldberg, M. E. (1991). Congruent representations of visual and somatosensory space in single neurons of monkey ventral intraparietal cortex (area VIP). In J. Paillard (Ed.), *Brain and space* (pp. 223–236). Oxford, NY: Oxford Press.

Duhamel, J. -R., Colby, C. L., & Goldberg, M. E. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science, 255,* 90–92.

Epstein, R. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in Cognitive Sciences, 12,* 388–396.

Fogassi, L., Gallese, V., di Pellegrino, G., Fadiga, L., Gentilucci, M., Luppino, G., & Rizzolatti, G. (1992). Space coding by premotor cortex. *Experimental Brain Research, 89,* 686–690.

Fuhs, M. C., & Touretzky, D. S. (2006). A spin glass model of path integration in rat medial entorhinal cortex. *Journal of Neuroscience, 19,* 4266–4276.

Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1991). Neuronal activity related to saccadic eye movements in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology, 65,* 1464–1483.

Fyhn, M., Molden, S., Witter, M. P., Moser, E. I., & Moser, M. B. (2004). Spatial representation in the entorhinal cortex. *Science, 305,* 1258–1264.

Goldman-Rakic, P. S. (1988). Topography of cognition: Parallel distributed networks in primate association cortex. *Annual Review of Neuroscience, 11,* 137–156.

Goodale, M. A., Meenan, J. P., Bulthoff, H. H., Nicolle, D. A., Murphy, K. J., & Racicot, C. I. (1994). Separate neural pathways for the visual analysis of object shape in perception and prehension. *Current Biology, 4,* 604–610.

Grefkes, C., & Fink, G. R. (2005). The functional organization of the intraparietal sulcus in humans and monkeys. *Journal of Anatomy, 207,* 3–17.

Hafting, T., Fyhn, M., Molden, S., Moser, M. B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature, 436,* 801–806.

Kakei, S., Hoffman, D. S., & Strick, P. L. (1999). Muscle and movement representations in the primary motor cortex. *Science, 285,* 2136–2139.

Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature Reviews Neuroscience, 12,* 217–230.

Ladavas, E. (1987). Is the hemispatial deficit produced by right parietal lobe damage associated with retinal or gravitational coordinates? *Brain, 110,* 167–180.

Lynch, J. C., & McLaren, J. W. (1989). Deficits of visual attention and saccadic eye movements after lesions of parietooccipital cortex in monkeys. *Journal of Neurophysiology, 61,* 74–90.

Marshall, J. C., & Halligan, P. W. (1993). Visuo-spatial neglect: A new copying text to assess perceptual parsing. *Journal of Neurology, 240,* 37–40.

Maunsell, J. H. (2004). Neuronal representations of cognitive state: Reward or attention? *Trends in Cognitive Sciences, 8,* 261–265.

McFie, J., & Zangwill, O. L. (1960). Visual-constructive disabilities associated with lesions of the left cerebral hemisphere. *Brain, 83,* 243–260.

McNaughton, B. L., Bataglia, F. P., Jensen, O., Moser, E. I., & Moser, M. B. (2006). Path integration and the neural basis of the "cognitive map. " *Nature Reviews Neuroscience, 7,* 663–678.

Moser, E. I., Kropff, E., & Moser, M. B. (2008). Place cells, grid cells and the brain's spatial representation system. *Annual Review of Neuroscience, 31,* 69–89.

Olson, C. R., & Gettner, S. N. (1995). Object-centered direction selectivity in the macaque supplementary eye field. *Science, 269,* 985–988.

Perenin, M. -T., & Vighetto, A. (1988). Optic ataxia: A specific disruption in visuomotor mechanisms. I. Different aspects of the deficit in reaching for objects. *Brain, 111,* 643–674.

Rao, S. C., Rainer, G., & Miller, E. K. (1997). Integration of what and where in the primate prefrontal cortex. *Science, 276,* 821–824.

Rizzolatti, G., & Luppino, G. (2001). The cortical motor system. *Neuron, 31,* 889–901.

Roesch, M. R., & Olson, C. R. (2003). Impact of expected reward on neuronal activity in prefrontal cortex, frontal and supplementary eye fields and premotor cortex. *Journal of Neurophysiology, 90,* 1766–1789.

Roesch, M. R., & Olson, C. R. (2004). Neuronal activity related to reward value and motivation in primate frontal cortex. *Science, 304,* 307–310.

Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience, 22,* 9475–9489.

Rondot, P., de Recondo, J., & Dumas, J. L. (1977). Visuomotor ataxia. *Brain, 100,* 355–376.

Sawaguchi, T., & Goldman-Rakic, P. S. (1994). The role of D1-dopamine receptor in working memory: Local injections of dopamine antagonists into the prefrontal cortex of rhesus monkeys performing an oculomotor delayed-response task. *Journal of Neurophysiology, 71,* 515–528.

Schwartz, A. B., Kettner, R. E., & Georgopoulos, A. P. (1988). Primate motor cortex and free arm movements to visual targets in three-dimensional space. I. Relations between single cell discharge and direction of movement. *Journal of Neuroscience, 8,* 2913–2927.

Sugrue, L. P., Corrado, G. S., & Newsome, W. T. (2004). Matching behavior and the representation of value in the parietal cortex. *Science, 304,* 1753–1754.

Teng, E., & Squire, L. R. (1999, August 12). Memory for places learned long ago is intact after hippocampal damage. *Nature, 400*(6745), 675–677.

Umeno, M. M., & Goldberg, M. E. (1997). Spatial processing in the monkey frontal eye field. I. Predictive visual responses. *Journal of Neurophysiology, 78,* 1373–1383.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.

Vaina, L. M. (1994). Functional segregation of color and motion processing in the human visual cortex: Clinical evidence. *Cerebral Cortex, 4,* 555–572.

Weiskrantz, L. (1996). Blindsight revisited. *Current Opinion in Neurobiology, 6,* 215–220.

Zola-Morgan, S., & Squire, L. R. (1993). Neuroanatomy of memory. *Annual Review of Neuroscience, 16,* 547–563.

Suggested Readings

Andersen, R. A., & Cui, H. (2009). Intention, action planning, and decision making in parietal-frontal circuits. *Neuron, 63,* 568–583.

Bisley, J. W., & Goldberg, M. E. (2010). Attention, intention, and priority in the parietal lobe. *Annual Review of Neuroscience, 33,* 1–21.

Gottlieb, J. (2007). From thought to action: The parietal cortex as a bridge between perception, action, and cognition. *Neuron, 53,* 9–16.

Hall, N. J., & Colby, C. L. (2011). Remapping for visual stability. *Philosophical Transactions of the Royal Society of London Series B - Biological Science, 366,* 528–539.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24,* 167–202.

Olson, C. R. (2003). Brain representation of object-centered space in monkeys and humans. *Annual Review of Neuroscience, 26,* 331–354.

Taube, J. S. (2009). Where am I and how will I get there from here? A role for posterior parietal cortex in the integration of spatial information and route planning. *Neurobiology of Learning and Memory, 91,* 186–196.

Carl R. Olson and Carol L. Colby

CHAPTER

46

Attention

INTRODUCTION

Intuition, together with cognitive and psychophysical experiments, shows that the brain is limited in the amount of information it can process at any moment in time. For instance, when people are asked to identify the objects of a briefly presented scene, they become less accurate as the number of objects increases. Similarly, when people concentrate on one demanding task (e.g., mental arithmetic) this typically comes at the expense of performance on other simultaneous activities (e.g., recalling a familiar tune). Limitations on the ability to simultaneously carry out multiple cognitive or perceptual tasks reflect the limited capacity of some stage or stages of sensory processing, decision making, or behavioral control. As a result of such computational bottlenecks, it is necessary to have neural mechanisms in place to ensure the selection of stimuli, or tasks, that are immediately relevant to behavior. "Attention" is a broad term denoting the mechanisms that mediate this selection.

In his *Principles of Psychology* published in 1890, William James wrote, "Everyone knows what attention is. It is the taking possession by the mind in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought.... " In modern neuroscience—specifically, neurophysiological research with animals—attention is defined with respect to the external world, as the focusing on one of many possible sensory stimuli. Despite their apparent discrepancies, these definitions have two features in common. They highlight the selective character of attention and highlight the cognitive nature of this selection: the fact that, whether directed without or within, attention guides selection of stimuli or thoughts.

Over the past several decades, research has concentrated on the relation between attention and sensory perception. Clearly, it is possible to focus on selected stimuli in any sensory modality—sights, sounds, smells, and touch. Most studies, however, have examined vision, the dominant sensory modality for humans and nonhuman primates. These experiments have investigated visual attention at multiple levels, ranging from behavior to the single neuron. This chapter outlines the view of attention that has emerged from these studies. A key observation is that attending to an object enhances the influence of that object over behavior and the extent to which it can be accurately perceived and reported. Conversely, withdrawing attention, either by force of the behavioral context or following certain brain lesions, can render observers practically blind to visual events. A second key point is that the attentional processes that underlie perceptual selection may also guide the voluntary eye movements with which foveate animals, such as monkeys and humans, scan the environment. A final point is that attention is closely coordinated with an ongoing task, gathering information that serves the current actions. In terms of neural organization, although some neural centers have been closely linked with the control of attention, attention appears to affect the activity of neurons at almost all levels of the visual system.

DEFINITIONS AND VARIETIES OF ATTENTION

Empirical studies have generally adopted two operational definitions of attention. Some studies, heralding back to early studies of Pavlovian learning, define an attended stimulus as one that preferentially controls behavior. Other, psychophysically based studies adopt a narrower definition, whereby an attended stimulus is one whose discriminability or detectability is modified (typically increased) during performance of a task. In general, these two definitions are well-aligned with each other. Consider for example coming to a busy intersection and

© 2013 Elsevier Inc. All rights reserved.

examining the traffic light before deciding what to do next. By the first definition we would say the light is attended since it controls one's future action (stop or proceed). By the second definition we would also say the light was attended, since the subject will be able to report the color of the light. We note, however, that while the equivalence of action-based and perception-based attention has been assumed in the field, the mechanisms of these processes may not be identical and are under active investigation.

Just as they have the capacity to flexibly link attention to action, individuals have great flexibility as to how, why, and to what precisely they attend. Attending to an object often is accompanied by overt orienting toward that object using the eyes and possibly also the head and body. When a person enters a room, it is natural to turn in that person's direction. However, one can also attend covertly to objects without looking directly at them. When driving, one can monitor a passing car while continuing to look straight ahead. Covert attention can improve peripheral visual acuity, thereby extending the functional field of view. Although one can direct attention without moving the eyes, the converse does not appear to be true: experiments show that covert attention must be deployed to an object before that object can be targeted with an eye movement. Indeed, saccades—rapid eye movements used most commonly for scanning the environment—may be considered a motor manifestation of visual attention, a relationship that is discussed in more detail later.

Certain external stimuli can summon attention in and of themselves. These can be physically salient objects, such as especially large, bright, or loud objects, or stimuli with special learned significance, such as one's own name or a mother's face. This mode of attentional engagement, known as *exogenous* or *stimulus driven*, ensures that salient and potentially important external events do not pass unnoticed even if they are not being actively sought out. However, purposeful behavior often requires that attention be directed voluntarily, or *endogenously*, based on internally defined goals and against potential external distractions. When reading, one purposefully directs one's attention from one word to the next, tuning out noise and other distractions. In natural behavior, endogenous and exogenous factors interact continuously to control the allocation of sensory processing (Egeth & Yantis, 1997).

Another important issue is *what* is attended. Attention can be directed to a location in space, regardless of what happens to be at that location. It can also be directed to a feature, as when one searches within a complex scene for an object of a particular shape or color. Attention can also select whole objects, as demonstrated by studies showing that when attention is directed to one feature of an object, the other features of the same object are selected automatically for visual processing and by studies finding that when attention is directed to one of two semitransparent objects, attending to a feature of one object impairs processing of the features of the other object. Finally, attention can select entire object categories in highly abstract ways (e.g., when you cross the street, you are looking for cars regardless of their different appearances). In each of these cases, attention has been found to facilitate processing of the attended location, feature, object, or object category, as assessed at the behavioral and neural levels (Peelen et al., 2009; Reynolds & Chelazzi, 2004).

Attention is thus highly flexible and can be deployed in a manner that best serves the organism's momentary behavioral goals: to locations, features, or objects, based on internal goals or the external environment, with or without accompanying orienting movements. It is important to keep in mind that although these phenomena can be isolated in laboratory experiments, all varieties of attention operate seamlessly during natural behavior.

NEGLECT SYNDROME: A DEFICIT OF SPATIAL ATTENTION

Studies of patients with brain lesions have identified regions of the brain that are involved in attention (see Fig. 46.1, which shows the major divisions of the human cortex). In the cortex, unilateral lesions in the parietal lobe, the frontal lobe, and the anterior cingulate cortex (Heilman, 1979; Vallar, 1993) in humans may cause a profound inability to attend to certain spatial regions, a syndrome known as spatial neglect. At the subcortical level, lesions of the basal ganglia or of the pulvinar thalamic nucleus, which is heavily connected with the parietal cortex, can also cause neglect. In severe cases, patients with neglect behave as if the world contralateral to their lesioned hemisphere (the contralesional world) has ceased to exist. For example, a patient with neglect following a right hemisphere lesion may fail to read from the left side of a book, may ignore the food on the left side of the plate, or may remain unaware of the numerals on the left side of a clock (Fig. 46.2). Neglect patients may also be reluctant to initiate movement in contralesional space, with or without external sensory stimulation. Because the critical feature causing an object to be ignored is its location, this type of neglect is thought of as primarily a disorder of spatial attention.

The behavioral impairments in neglect cannot be explained by simple sensory or motor deficits (Mesulam, 1999). Neglect patients have normal vision in the contralesional visual field once their attention has been directed there, and they have no hemiparesis that could account for their reluctance to move. In addition, the extent of their difficulty is not immutable, as a sensory or motor deficit

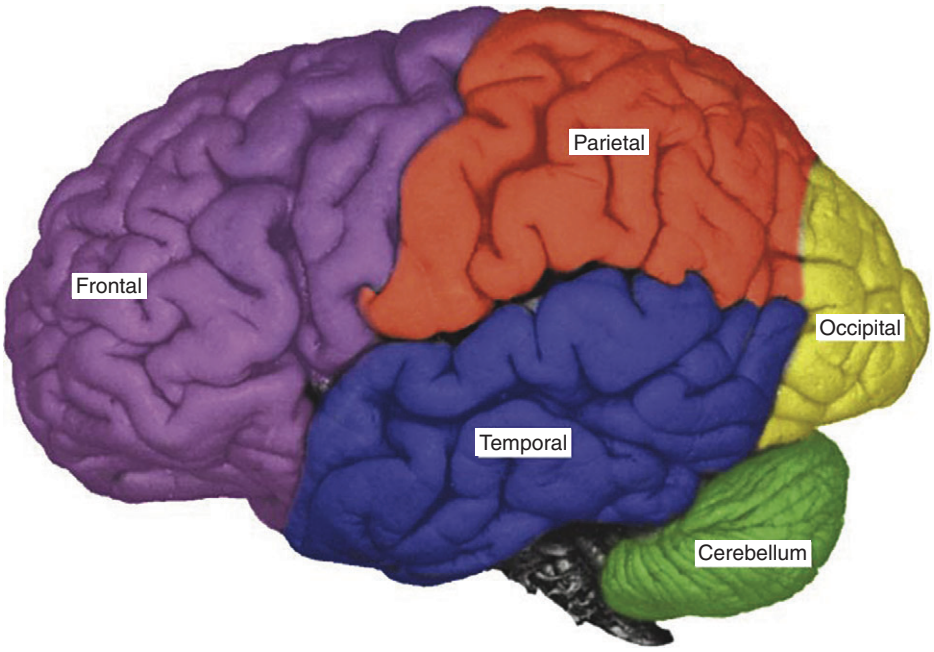


FIGURE 46.1 Lateral view of a human brain. Frontal (purple), parietal (orange), temporal (blue), and occipital (yellow) lobes are outlined.

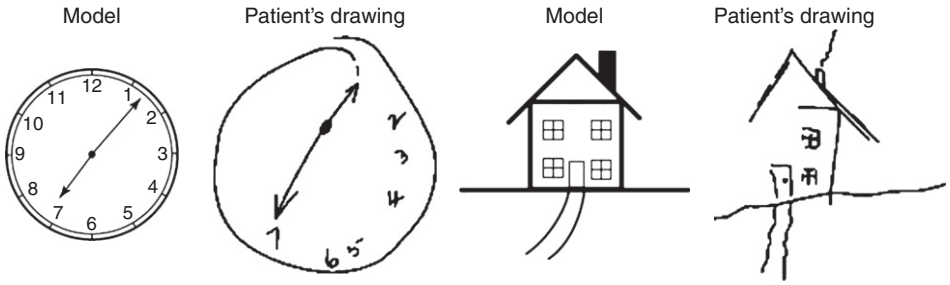


FIGURE 46.2 Two drawings that were made by a patient with spatial neglect. The patient was asked to copy the two models (clock, house). In each case, the copies exclude important elements that appeared on the left side of the model, indicating that the patient was unable to process information about the left side of the model.

would be, but is strongly affected by the goals, expectations, and motivational state of the patient. Thus, neglect represents a failure to select, in some circumstances, the appropriate portions of a sensory representation. Neglect operates on high-level representations that differ considerably from the raw sensory input. For example, visual neglect is not confined only to retinal visual coordinates (affecting only objects in the contralesional visual field) but can affect objects in the contralesional space relative to the patient's body, relative to the patient's gaze, or relative to an external object or scene. Thus, neglect affects a visual representation that incorporates information about the position of the body with high-level visual information. Neglect affects not only the perception of the present sensory environment, but also the processing of memorized or imagined scenes.

A related but milder attentional deficit, known as extinction, can follow more limited brain lesions or can occur during recovery from the acute phase of neglect. In visual extinction, patients are able to orient toward a contralesional stimulus presented in isolation, but fail to notice it if the same stimulus is presented simultaneously with an ipsilateral distracter. Extinction-like deficits are demonstrated readily with the spatial cuing task (Posner, Walker, Friedrich, & Rafal, 1984). In these tasks, subjects are cued to attend to the left or right visual field. Unimpaired observers are better able to process information at the cued location, but can also detect stimuli at an uncued location. Compared to unimpaired observers, neglect patients are slowed only mildly in their ability to detect targets in the contralesional visual field following valid symbolic or peripheral cues in the same field. However, they have enormously prolonged latencies for detecting contralateral targets if their attention had been misdirected to the ipsilesional field by means of an invalid cue. No impairment is seen when patients are cued to the contralesional field but asked to detect an ipsilesional target. Extinction-like phenomena also have been demonstrated in search tasks in which patients are asked to find a target object located at various locations in a complex

scene. The time needed to find a target in contralesional space increases in proportion with the number and salience of ipsilesional distracters.

ATTENTION AFFECTS NEURAL ACTIVITY IN THE HUMAN VISUAL CORTEX IN THE PRESENCE AND ABSENCE OF VISUAL STIMULATION

Human functional brain imaging studies are in agreement with both single-cell physiology and neuropsychological studies in identifying areas in the parietal, frontal, and cingulate cortices as being especially active in relation to spatially directed attention. When observers attend to a location in space in anticipation of the appearance of a stimulus, this is accompanied by elevated levels of activity in a fronto-parietal network consisting of the superior parietal lobule (SPL), the frontal eye field (FEF), and the supplementary eye field (SEF) extending into the anterior cingulate cortex (Fig. 46.3). This elevation in activity persists throughout the expectation period and when stimuli later appear at the attended location (Fig. 46.4B). It is important to note that the fronto-parietal attentional control network is not limited to controlling spatial attention. It is also activated when subjects select nonspatial information. Object- or feature-based attention engages the fronto-parietal attention network as well, indicating a general domain-independent mechanism of attentional control (Yantis & Serences, 2003).

Frontal and parietal areas have been associated with the control of spatial attention, but they are not usually considered crucial for visual processing or object recognition. There is converging evidence from event-related potential (ERP) and functional imaging studies that feedback signals from the fronto-parietal areas that control the allocation of attention can affect the neural processing of visual information in the human visual cortex. In a typical experiment, identical visual stimuli are presented simultaneously to corresponding peripheral field locations to the right and to the left of a central fixation point. Subjects are instructed to direct attention covertly to the right or the left by a symbolic cue presented at the fixation point and to detect the occurrence of a visual stimulus or, in some cases, discriminate a feature of the attended stimulus. Directing attention to the left hemifield increases stimulus-evoked activity in the extrastriate visual cortex of the right hemisphere, whereas directing attention to the right hemifield increases activity in the extrastriate

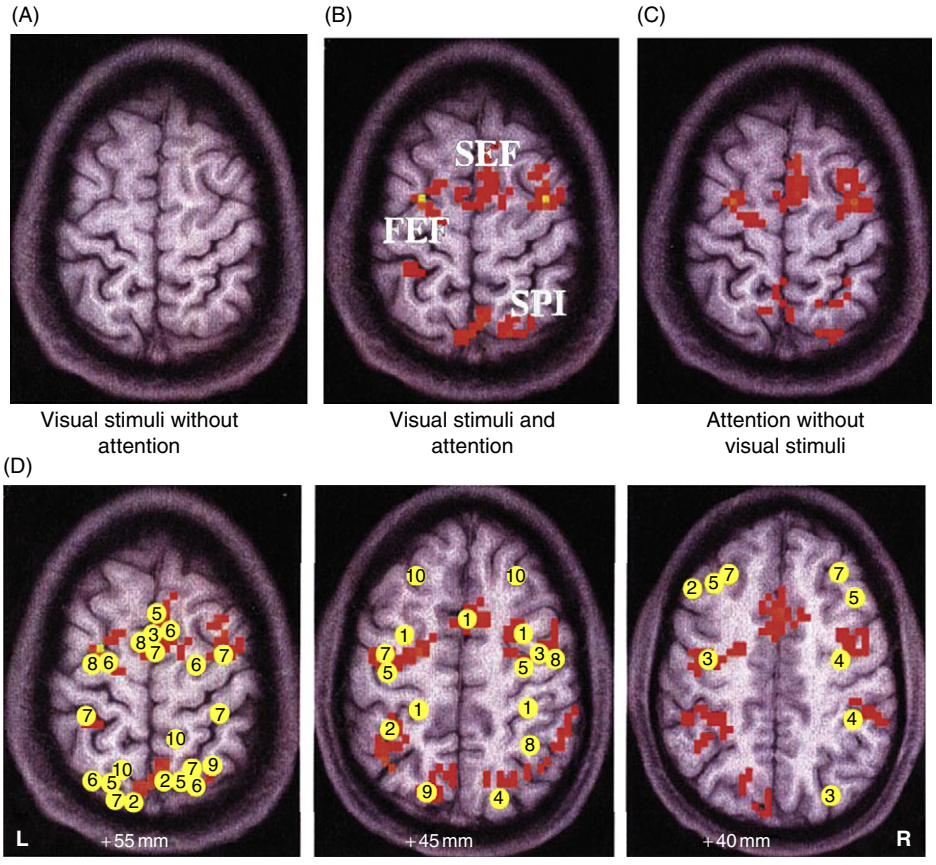


FIGURE 46.3 Regions in human brain activated by attention and regions associated with neglect. (A) Visual stimulation did not activate the frontal or parietal cortex reliably when attention was directed elsewhere in the visual field. (B) When the subject directed attention to a peripheral target location and performed an object discrimination task, a distributed fronto-parietal network was activated, including the SEF, the FEF, and the SPL. (C) The same network of frontal and parietal areas was activated when the subject directed attention to the peripheral target location in the expectation of the stimulus onset—in the absence of any visual input whatsoever. This activity therefore may not reflect attentional modulation of visually evoked responses, but rather attentional control operations themselves. (D) Meta-analysis of studies investigating the spatial attention network. Axial slices at different Talairach planes are indicated. Talairach (peak) coordinates of activated areas in the parietal and frontal cortex from several studies are indicated (For references, see Kastner & Ungerleider, 2000). R, right hemisphere; L, left hemisphere.

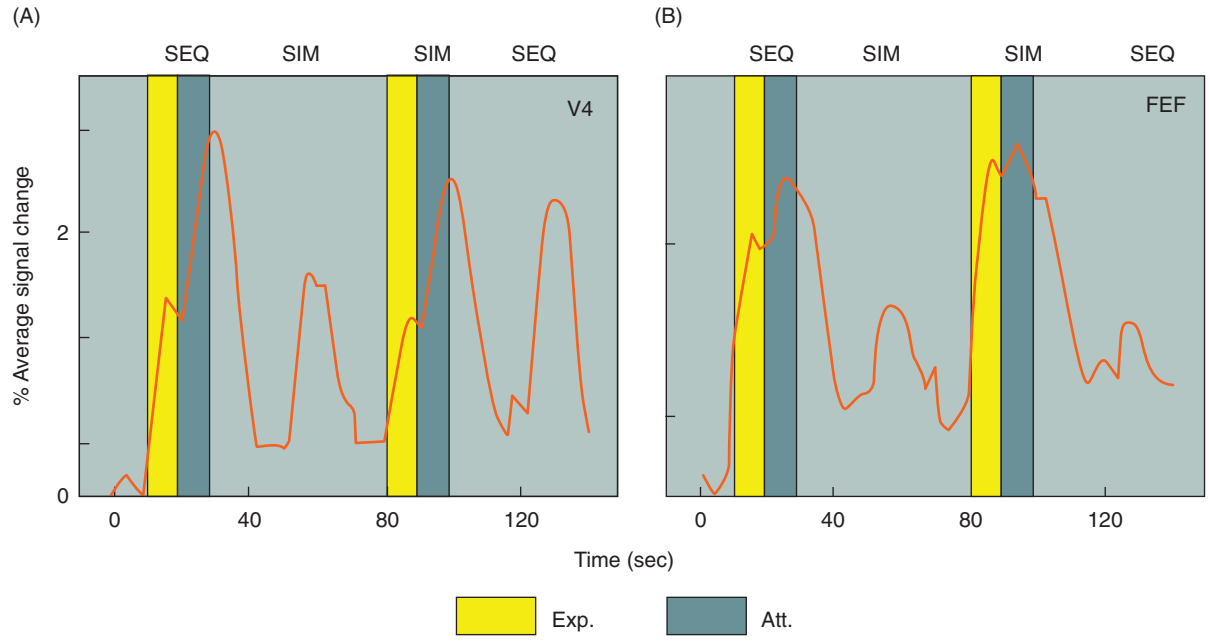


FIGURE 46.4 Directed attention in humans with and without visual stimulation. (A) Time series of fMRI signals in V4. Directing attention to a peripheral target location in the absence of visual stimulation led to an increase of baseline activity (textured blocks), which was followed by a further increase after the onset of stimuli (gray-shaded blocks). Baseline increases were found in both the striate and the extrastriate visual cortex. (B) Time series of fMRI signals in FEF. Directing attention to the peripheral target location in the absence of visual stimulation led to a stronger increase in baseline activity than in the visual cortex; the further increase of activity after the onset of stimuli was not significant. Sustained activity was seen in a distributed network of areas outside the visual cortex, including SPL, FEF, and SEF, suggesting that these areas may provide the source for the attentional top-down signals seen in visual cortex. *Adapted from Kastner et al. (1999).*

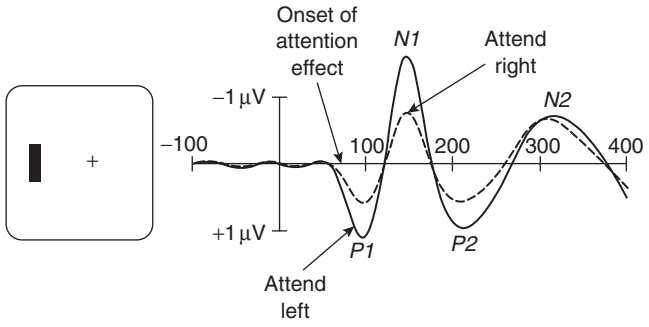


FIGURE 46.5 Attentional modulation of event related potentials (ERPs). In a typical experiment, subjects fixate a central cross and attend either to the left or right visual field. Stimuli are then presented to the left and right visual fields in a rapid sequence. This idealized example illustrated the common finding that the average ERP elicited by an attended visual field stimulus contains larger P1 and N1 components.

visual cortex of the left hemisphere (Heinze et al., 1994) (Fig. 46.5). Thus, responses to stimuli are enhanced on the side of the extrastriate cortex that contains representations of the attended hemifield. As is the case in the fronto-parietal areas, this increase in response is observed during the period prior to appearance of a stimulus, reflecting the effects of feedback signals from the attentional control areas.

In one imaging study (Kastner, Pinsk, De Weerd, Desimone, & Ungerleider, 1999), subjects were cued to direct attention covertly to a target location in the periphery of the visual field and to anticipate the onset of visual stimuli. Visual stimuli occurred with a delay of several seconds. Neural activity associated with spatially directed attention in the absence and in the presence of visual stimulation could therefore be distinguished. As was the case in the fronto-parietal control network, fMRI signals increased when attention was directed to the target location and before any visual stimulus was present on the screen. This increase in baseline activity was followed by an additional increase when stimuli appeared at the attended location (Fig. 46.4A), reflecting attention-dependent increases in the stimulus-evoked response itself. These studies and related single-unit studies in the monkey show that spatial attention effects are topographically organized and retinotopically specific, resulting in increased activity among neurons coding the attended location. Taken together, these findings suggest that an important way that spatial attention modulates visual processing is by amplifying the neural signals for stimuli at an attended location.

Functional brain imaging studies and single-unit recording studies in nonhuman primates have found evidence that nonspatial attention also modulates sensory cortices. One such study found that selective attention to shape, color, or speed enhanced activity in the regions of the extrastriate visual cortex that selectively process these same attributes. Attention to shape and color led to response enhancement in regions of the posterior portion of the fusiform gyrus, including area V4. Attention to speed led to response enhancement in areas MT/MST (Corbetta, Miezin, Dobmeyer, Shulman, & Petersen, 1991). In other studies, attention to faces or houses led to response enhancement in areas of the midanterior portion of the fusiform gyrus, areas responsive to the processing of faces and objects. These results support the idea that selective attention to a particular stimulus attribute modulates neural activity in those extrastriate areas that preferentially process the selected attribute.

SINGLE UNIT RECORDING STUDIES IN NONHUMAN PRIMATES PROVIDE CONVERGENT EVIDENCE FOR A FRONTO-PARIETAL ATTENTIONAL CONTROL SYSTEM

Studies in monkeys have supported the conclusion that frontal and parietal cortices play a role in controlling the allocation of attention (Bushnell et al., 1981). These studies point to two interconnected areas, one in the parietal cortex, the lateral intraparietal area (LIP), and one in the frontal lobe, the frontal eye field (FEF), as being important for spatial attention. Many neurons in FEF and LIP have spatially selective visual receptive fields, meaning that they respond to visual stimuli that activate a limited region of the retina. These receptive fields are retinotopic (they are linked to the retina and move relative to the external world every time the eye moves) and mostly contralateral, so that the FEF and LIP in one side of the brain represent the locations of objects in the contralateral portion of the visual field. Thus, the neurons provide a map of stimulus location in retinotopic coordinates.

Unlike the maps found in early visual areas (including the retina itself), the maps in LIP and FEF do not indicate the mere presence of an object in the world but respond selectively to objects that are likely to be attended, whether by virtue of their physical properties (conspicuity) or behavioral relevance. This was shown in a task where monkeys made eye movements (saccades) across a stable visual background, such that inconspicuous objects moved in and out of a neuron's receptive field (Gottlieb et al., 1998). LIP neurons responded little if a stable object entered their receptive field (Fig. 46.6A, center). However, they responded strongly if the identical object was presented with an abrupt onset that rendered it salient. Neural responses are strong if the object appears directly in the receptive field as is commonly the case in physiological experiments, but also if the object appears elsewhere and enters the receptive field a fraction of a second later by virtue of the subject's eye movement (A, right). An additional experiment showed that neurons also respond to stimuli that are not physically salient but become relevant, for example, by being selected as the target of the next saccade (B). These findings suggest that LIP integrates bottom-up and top-down factors relevant for attention and encode the priority of objects in the world. A similar proposal has been advanced for the FEF, and may also hold for area 7a, a parietal area adjacent to LIP (Constantinidis & Steinmetz, 2001).

A subsequent experiment linked the selective responses in LIP with attention measured behaviorally, as a transient enhancement of perceptual sensitivity (Bisley & Goldberg, 2003). In this experiment monkeys were shown a brief visual stimulus whose location they had to remember for a later saccade. During the memory period, a salient but irrelevant distractor flashed in the visual field. The authors measured the locus of attention at different times and locations by using a secondary contrast discrimination task. Even though the eyes remained still during the memory period, attention shifted—moving to the distractor location immediately after the flash and back to the saccade target later during the delay period. These dynamics were encoded in the neural activity in LIP: attention was pinned at the distractor location for as long as the LIP response to the distractor exceeded that to the target, and flipped to the saccade goal when the balance of activity shifted in LIP. Thus, the selective visual responses in LIP are driven both by salient and task-relevant stimuli and reflect the momentary locus of attention. Similar correspondences are found for FEF neurons in visual search tasks where subjects must direct attention to one of many stimuli in a display (Thompson & Bichot, 2005).

Direct evidence that LIP and FEF are involved in attention comes from experiments that chemically inactivate or microstimulate these areas and measure the effects of these manipulations on attention. Inactivation of the LIP or FEF using microinjections of muscimol (a GABA-A receptor agonist that temporarily silences spiking activity at the injection site) produces deficits in target selection during visual search. In the FEF, subthreshold microstimulation biases covert attention toward the location represented by the neurons at the stimulation site. Microstimulation of the FEF also modulates neural activity at corresponding locations in area V4, consistent with the idea that FEF neurons send top-down feedback to earlier visual areas.

In addition to modulating visual processing, LIP and FEF influence saccade motor planning through their

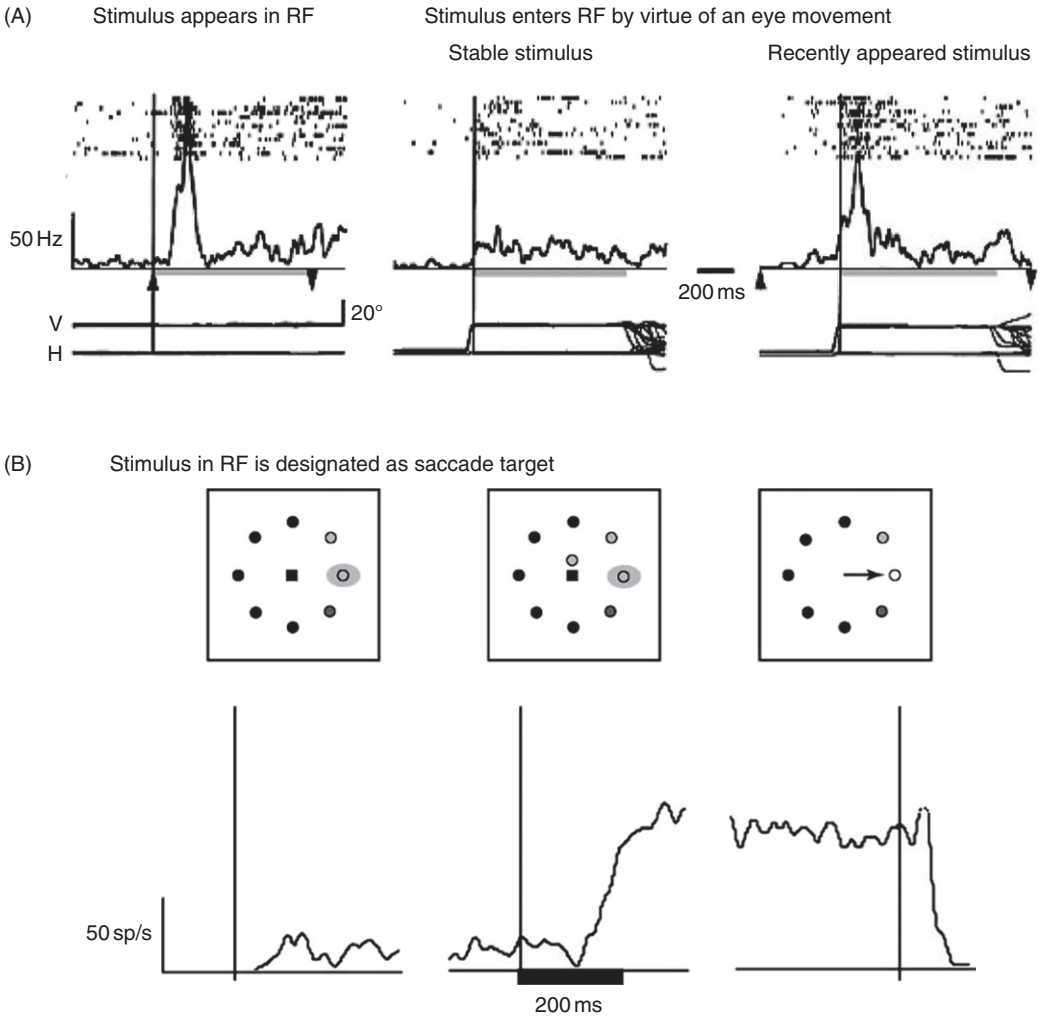


FIGURE 46.6 LIP neurons encode a spatial salience representation. (A) LIP neurons are sensitive to the physical conspicuity of objects in the receptive field. An LIP neuron responds strongly when a visual stimulus is flashed in its receptive field (left). The same neuron does not respond if the physically identical stimulus is stable in the world and enters the receptive field by virtue of the monkeys' eye movement (center). The neuron again responds if the stimulus entering the receptive field is rendered salient by flashing it on and off on each trial (before it enters the receptive field). In each panel, rasters show the times of individual action potentials relative to stimulus onset (left) or the eye movement that brings the stimulus in the receptive field (center and right). Traces underneath the rasters show average firing rate. Bottom traces show the horizontal and vertical eye position. The rapid eye movement (saccade) that brings the stimulus in the receptive field is seen as the abrupt deflection in eye position traces. (B) LIP neurons are sensitive to the behavioral relevance of objects in the receptive field. Traces show average firing rate of an LIP neuron when the monkey is cued to make a saccade to a stimulus in the receptive field. In the left panels the monkey is fixating and a stable stimulus is in the receptive field. The neuron has low firing rate. In the center panel the monkey receives a cue instructing him to make a saccade to the receptive field. The neuron gradually increases its response. In the right panel the monkey executes the instructed saccade. The neural response remains high until after the eye movement.

descending connections to the superior colliculus and brainstem oculomotor centers. Microstimulation and neural recordings suggest that these areas form a continuum of visuo-motor transformation, with LIP having primarily a visual role and the FEF being related to both visual selection and oculomotor output. Consistent with this idea, thresholds for electrically evoking saccades are high in LIP but much lower in the FEF. This suggests that LIP is more closely related to visual selection whereas FEF has a stronger influence on eye movement output. Similarly, neural recording studies show that LIP neurons are more closely related to visual selection whereas FEF encodes both visual selection and oculomotor output. In these studies monkeys are trained to make eye movements toward blank spatial locations, which may be *opposite* a salient visual cue, dissociating visual and motor selection. Such tasks show that LIP neurons predominantly encode stimulus location but provide scant motor information, responding only weakly before nonvisually guided saccades. The FEF contains a population of neurons that are predominantly sensitive to visual cues, but also contains neurons with mixed visuo-motor properties

and a small class of neurons—"movement neurons"—that are tightly coupled with saccade output. These neurons respond to saccades but not to simple visual inputs and ramp their responses to a constant level (threshold) prior to a saccade, suggesting that they reflect a fixed saccade threshold. These findings suggest that the brain implements a continuous transformation starting with a stage of visual selection mediated by visual and visuo-motor neurons in LIP and FEF, and culminating in motor selection mediated by movement neurons in FEF.

The properties of LIP and FEF neurons described above have lent support to the idea that the attentional system uses a central stage of visual selection—a "salience" or "priority map" that indicates a common currency for directing attention (Fig 46.7). Computational vision models show that systems that contain such a centralized architecture are more efficient than those implementing a more distributed selection in feature-specific visual areas. In the latter case attention would be allocated through competition between saliency signals in individual feature domains: for example, a color-selective area may tend to direct attention to a location of a color pop-out, competing with a motion selective area that attracts attention to a location with salient motion. By allowing these separate feature signals to converge to a common stage the brain can more efficiently coordinate visual selection as well as coordinate selection with action.

Experiments in recent years have begun to address a related and equally important question: namely, how are attention and eye movements coordinated with an ongoing task? How do neurons in FEF and LIP "know" which of the many available stimuli is related to one's current actions? The answer that is emerging from recent work suggests that this learning may be mediated by feedback regarding actions and reward to selective visual representations.

Studies from multiple laboratories show that visual selection response in LIP is not stereotyped but is shaped by the task associations of the attended stimulus. Among the factors modulating LIP activity are the shape or color of a search target, the statistical relationship between a target and salient stimulus, and the category of a visual cue in a categorization task. Figure 46.8A shows an example of modulatory feedback related to the motor response with which a monkey reported a perceptual discrimination (Oristaglio, Schneider, Balan, & Gottlieb, 2006). Monkeys were trained to find a cue in an array of distractors, and, without shifting gaze to the cue, report the

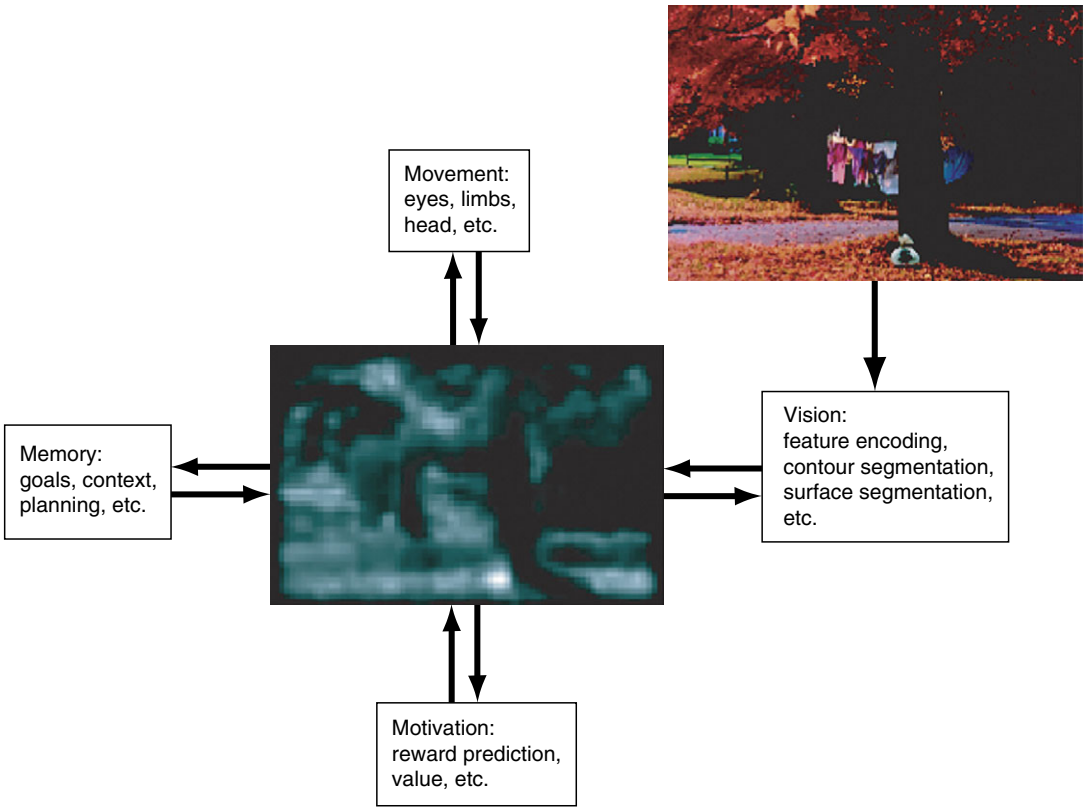


FIGURE 46.7 A salience representation can be viewed as an intermediate representation that interacts with multiple behavioral systems (visual, motor, cognitive, and motivational) and fashions a unified signal of "salience" based on multiple task demands. By virtue of feedback connections to these systems the salience map can help coordinate output processing in multiple "task-relevant areas."

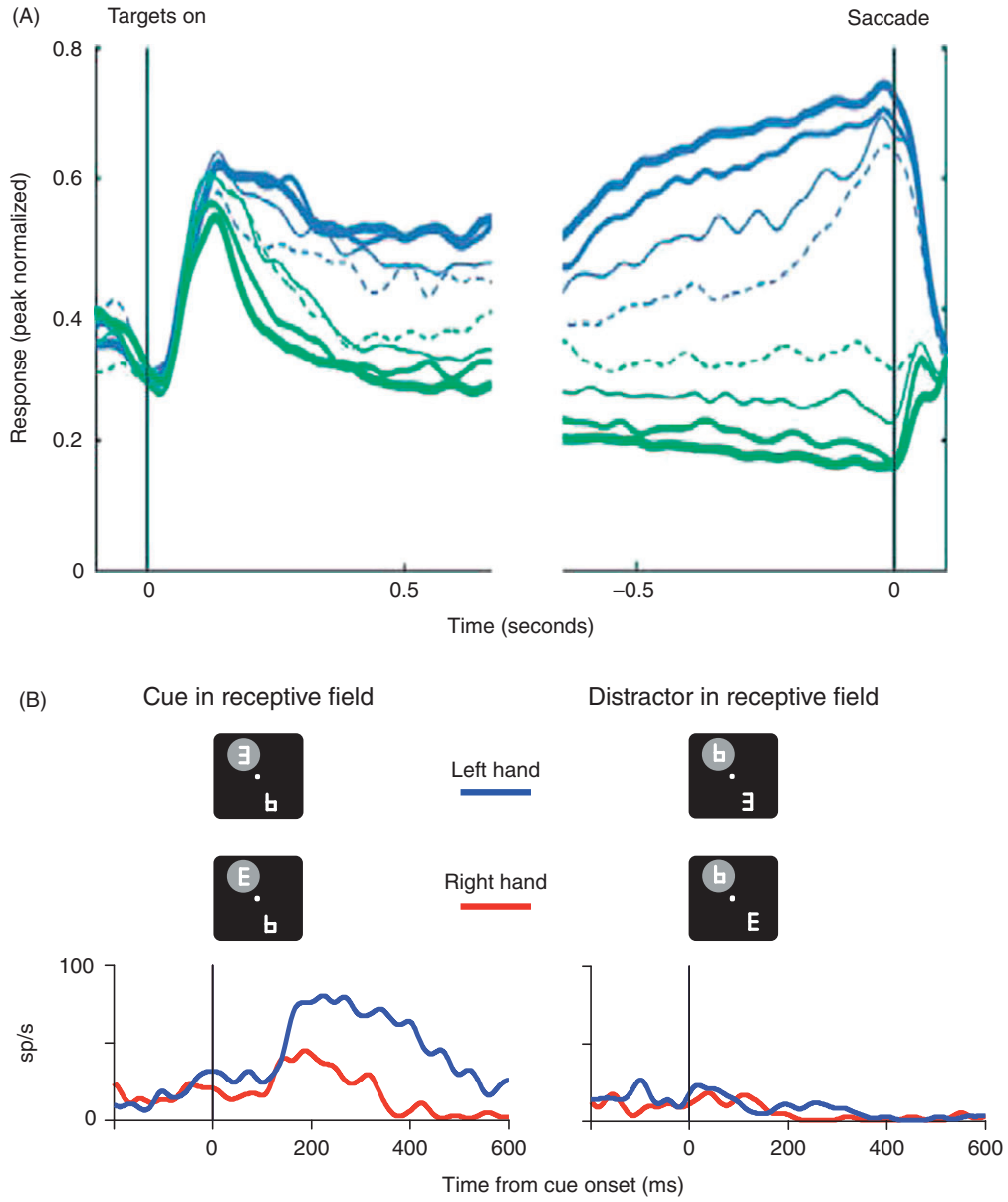


FIGURE 46.8 (A) LIP responses are modulated by reward probability. Monkeys performed a dynamic foraging task in which they tracked the changing reward values of each of two saccade targets. Traces represent population responses for neurons with significant effects of reward probability. Blue traces represent saccades toward the receptive field, and green traces represent saccades opposite the receptive field. Traces are further subdivided according to local fractional income (reward probability during the past few trials): solid thick lines, 0.75–1.0; solid medium lines, 0.5–0.75; solid thin line, 0.25–0.5; dotted thin lines, 0–0.25. Activity for saccades toward the receptive field increased, and that for saccades away decreased as function of local fractional income, resulting in more reliable spatial selectivity (difference between the two saccade directions) with increasing reward probability. (B) LIP neurons are modulated by limb motor planning. Monkeys viewed a display containing a cue (a letter "E") and several distracters. Without moving gaze from straight ahead, they reported the orientation of the cue by releasing one of two bars grasped with their hands. The neuron illustrated here responded much more strongly when the cue rather than a distracter appeared in its receptive field (left versus right panels). This cue-related response was modulated by the manual release: the neuron responded much more when the monkey released the left than when she released the right bar (blue versus red traces). Limb modulations did not always accompany a bar release, but were found only when the attended cue was in the receptive field. Bar-release latencies were on the order of 400–500 ms.

orientation of the cue with a manual release. Monkeys reported whether the cue was right- or left-facing by releasing a bar grasped with, respectively, their right or left paw. LIP neurons responded much more strongly to a cue, relative to a distractor, in their receptive fields (Fig. 46.8A, left vs. right panels). Thus, neurons encoded the location of the cue, as would be expected based on previous work. Unexpectedly, however, the cue selection response was modulated by the monkey's choice of limb. The neuron illustrated in Figure 46.8A responded more

strongly to the cue in its receptive field if the monkey released the left bar than if she released the right bar, while other neurons showed the opposite preference, responding more for right than for left bar release. This finding was surprising because the motor response was nontargeting—not directed to a particular spatial location—and was performed outside the monkey's field of view. Thus, the visual selection signal in LIP was modulated by a nonspatial variable, the limb that was associated with the selected visual cue. While the significance of this nonspatial signal is not yet fully understood, an intriguing possibility is that it may represent feedback through which the brain learns the relation of the cue to a particular motor output and hence, its "relevance" in the behavioral task.

A particularly significant source of feedback carries information about expected reward. An illustration of reward effects in LIP comes from an experiment where monkeys were free to choose between two saccade directions (either toward or away from the receptive field), and the probability that any one direction would be rewarded was systematically varied (Fig. 46.8B; Sugrue, Corrado, & Newsome, 2004). Monkeys apportioned their choices in relation to reward probability, tending to choose the target with the higher probability of reward. LIP responses to the selected target increased as a function of reward probability (blue traces). A later experiment showed that reward expectancy modulates responses to salient stimuli even if these are not saccade targets, consistent with the predominantly visual nature of the selective response in LIP (Peck, Jangraw, Suzuki, Efem, & Gottlieb, 2009). This suggests that priority maps may associate stimuli with rewards, possibly guiding attentional allocation so as to optimize performance in a task.

ATTENTION INCREASES NEURONAL RESPONSES AND BOOSTS THE CLARITY OF SIGNALS GENERATED BY NEURONS IN PARTS OF THE VISUAL SYSTEM DEVOTED TO PROCESSING INFORMATION ABOUT OBJECTS

Consistent with the previously mentioned functional imaging studies, single-unit recording studies in the visual cortices have found that when attention is directed to a single stimulus in the receptive field, there is often an increase in the firing rates of neurons that respond to the attended stimulus (Reynolds & Chelazzi, 2004). This is illustrated in Figure 46.9A. On the left is the average firing rate of an individual neuron recorded in macaque Area V4. The neuron's response was higher when the monkey attended to the stimulus within the neuron's receptive field (red line) than when attention was directed away to a distant stimulus (blue). This attention-dependent increase in mean firing rate is typically observed when attention is directed toward a single neuron within the receptive field, as shown in the middle figure of A, which shows the distribution of attentional modulation across the population, with positive values corresponding to neurons exhibiting attention-dependent increases in mean response and negative values corresponding to reductions. Gray bars indicate neurons whose mean firing rates were not significantly modulated by attention. In addition to this modulation of mean firing rate, attention also reduces the variability of the neuronal response over time (Mitchell, Sundberg, & Reynolds, 2007). This is illustrated in the rightmost figure of A. Here, positive values correspond to neurons whose responses became more variable ("noisier") when attention was directed toward the stimulus within the receptive field, and negative values, to neurons whose responses became less variable.

How much does this reduction in variability translate into improved signaling across a neuronal population? This depends on whether the variability of neuronal responses that is reduced by attention is correlated across the neuronal population. Variability that is *uncorrelated* can be averaged away when the population responses are pooled by a downstream neuron, but *correlated* variability cannot be averaged out. Cortical neuronal responses are typically somewhat correlated with one another (Harris & Thiele, 2011), as shown in Figure 46.9B. Here, neuronal responses were recorded simultaneously from two neurons, and the spiking responses were summed over brief time intervals. The resulting spike counts for a pair of V4 neurons are shown in the blue panel on the left. The responses of the two neurons were correlated, with stronger responses in Neuron 1 typically being accompanied with stronger responses in Neuron 2. This correlation was substantially reduced when attention was directed toward the stimulus that fell within the two neurons' receptive fields, as illustrated in the red panel, center. Across the population, this firing rate correlation was diminished by approximately one-half, as shown in the panel on the right (Mitchell, Sundberg, & Reynolds, 2009).

So attention modulates the mean firing rates of neurons (a change in their *gain*) and simultaneously decorrelates the moment-to-moment variability in individual neurons' firing rates. Theoretical calculations enable us to calculate the improvement in the signal quality (*signal-to-noise ratio*) resulting from these increases in the gain and reductions in correlation of the afferent responses (Zohary, Shadlen, & Newsome, 1994). This is illustrated in the left figure of Figure 46.9C, which shows the signal-to-noise ratio of the pooled signal as a function of the number of neurons in the pool. If the interneuronal variability is uncorrelated, the signal-to-noise ratio increases linearly with the logarithm of the number of neurons in the pool (straight line). If neuronal variability is correlated across neurons, this imposes an upper bound on the signal-to-noise ratio of

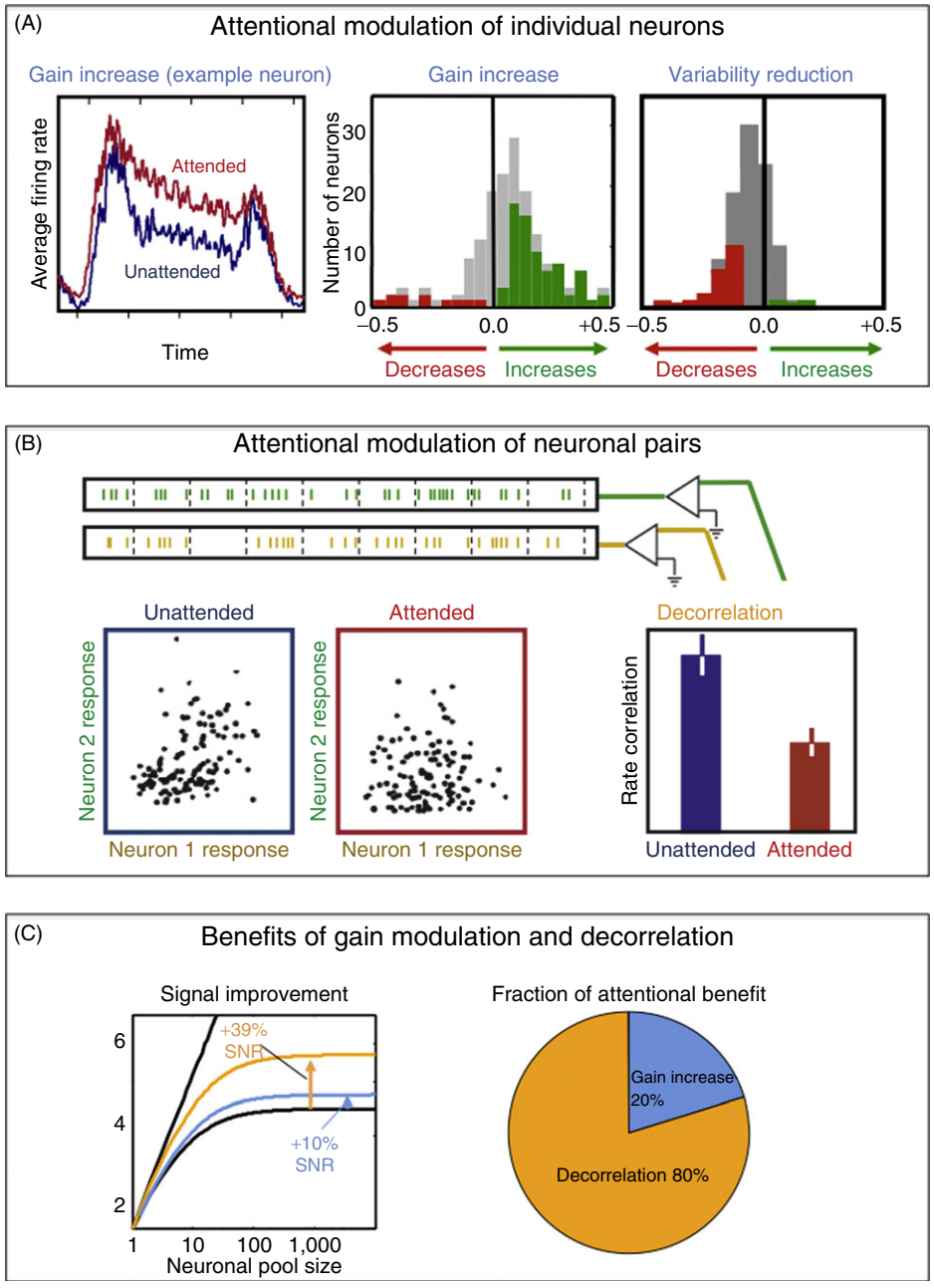


FIGURE 46.9 Attention increases the mean firing rates of individual neurons and the degree to which their firing rates covary over time. (A) Attention-dependent increases in firing rate. (B) Attention-dependent reductions in response covariation. (C) Under the assumption that downstream neurons pool the signals of the afferent neuronal population, observed levels of attention-dependent decorrelation of low-frequency (<10 Hz) fluctuations in firing rate result in increases in signal quality four times as large as the improvement attributable to increases in firing rate.

the pooled signal. This is shown in the saturating black curve, which was calculated based on correlations actually observed in macaque V4 neurons when attention was directed away from the stimulus. The blue line shows the modest increase in asymptotic signal quality that results from the measured attention-dependent increase in gain. The yellow line shows the comparatively larger increase in signal quality that results from observed attention-dependent reductions in low-frequency response correlations (Mitchell et al., 2009). The improvements due to gain and decorrelation are shown in the figure to the right. Decorrelation accounts for the lion's share of attention-dependent improvement in signal quality (Cohen & Maunsell, 2009; Mitchell et al., 2009).

Microstimulation studies support the conclusion that attention-dependent changes in firing rate within the visual cortices result from feedback signals that are generated in the fronto-parietal attentional control network. Moore and Armstrong (2003) measured the locations in FEF where stimulating current would move the eyes, the "movement field" of the stimulation site. They then reduced the stimulating current to a level too low to evoke an eye movement, and observed the effect of this stimulation on the responses of neurons in visual area V4. They

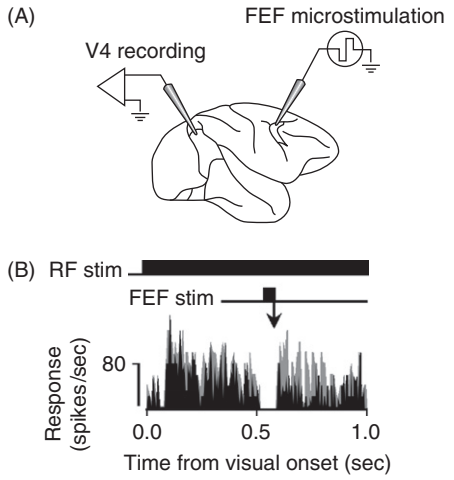


FIGURE 46.10 Electrical stimulation of FEF increases neuronal responsiveness in V4. (A) Current was injected into FEF while neuronal activity was recorded from V4. (B) The visual stimulus appeared in the receptive field for one second (RF stim). 500 ms after the onset of the visual stimulus, a low level of current was applied to a site in FEF for 50 ms (FEF stim). The response of a single V4 neuron with (gray) and without (black) FEF microstimulation appears below. The apparent gap in response reflects the brief period during which V4 recording was paused while current was injected into FEF, when, for technical reasons, the stimulating current interfered with the ability to record action potentials. Following stimulation, the V4 neuron had elevated responses on trials when FEF was electrically stimulated, as compared to trials without FEF stimulation.

found that FEF stimulation caused V4 neurons to respond more robustly to a stimulus that appeared at the movement field within the V4 neuron's receptive field (Fig. 46.10A). Figure 46.10B shows responses of a single V4 neuron with and without stimulation. The time course of stimulus presentation (RF stim) and current injection (FEF stim) are illustrated at the top of the panel. The neuronal responses in the two conditions are indicated at the bottom of the figure. Stimulation was injected 500 ms after the appearance of the stimulus, at which point the stimulus-evoked response had begun to diminish in strength. The average response on microstimulation trials (gray) was elevated following electrical stimulation, relative to nonstimulation trials.

ATTENTION MODULATES NEURAL RESPONSES IN THE LATERAL GENICULATE NUCLEUS

The lateral geniculate nucleus (LGN) is the thalamic station in the retinocortical projection and traditionally has been viewed as the gateway to the visual cortex. In addition to retinal afferents, the LGN receives input from multiple sources including striate cortex, the thalamic reticular nucleus (TRN), and the brainstem. The LGN therefore represents the first stage in the visual pathway at which cortical top-down feedback signals could affect information processing.

It has proven difficult to study attentional response modulation in the LGN using single-cell physiology due to the small RF sizes of LGN neurons and the possible confound of small eye movements. Initially, several single-cell physiology studies failed to demonstrate attentional modulation in the LGN, supporting a notion that selective attention affects neural processing only at the cortical level. This notion was first revisited using fMRI in humans (O'Connor, Fukui, Pinsk, & Kastner, 2002). As in visual cortex, different modulatory effects of attention were found. Neural responses to attended visual stimuli were enhanced relative to the same stimuli when unattended. As in cortex, this effect of attentional response enhancement was shown to be spatially specific in the LGN. And directing attention to a location in the absence of visual stimulation and in anticipation of the stimulus onset increased neural baseline activity. Together, these studies indicate that the LGN appears to be the first stage in the processing of visual information that is modulated by attentional top-down signals.

These studies were suited to compare the magnitude of the attention effects across the visual system. The magnitude of all attention effects increased from early to more advanced processing levels along both the ventral and dorsal pathways of visual cortex. This is consistent with the idea that attention operates through top-down signals that are transmitted via cortico-cortical feedback connections in a hierarchical fashion. This idea is supported by single-cell recording studies, which have shown that attention effects in area TE of inferior temporal cortex have a latency of approximately 150 ms, whereas attention effects in V1 have a longer latency of approximately 230 ms. According to this account, one would predict smaller attention effects in the LGN than in striate cortex. Surprisingly, it was found that all attention effects tended to be larger in the LGN than in striate cortex. This finding suggests that attentional response modulation in the LGN is unlikely to be due solely to corticothalamic feedback from striate cortex, but may be further influenced by additional sources of input. In addition to corticothalamic feedback projections from V1, which comprise about 30% of its modulatory input, the LGN receives another 30% of modulatory inputs each from the TRN and the parabrachial nucleus of the brain stem. This view has been corroborated by a recent physiology study (McAlonan, Cavanaugh, & Wurtz, 2008) that demonstrated concurrent response modulation in parvo- and magnocellular LGN and the TRN. While the response amplitudes evoked by an attended stimulus increased relative to those evoked by unattended stimuli in the LGN, they decreased in the TRN and also

preceded the LGN responses, suggesting a release of inhibitory control of the TRN over the LGN. These effects occurred in the first 100 ms after stimulus onset. A second effect of attention on the LGN manifested 200 ms after stimulus onset and was independent of the TRN. It is conceivable that this late response modulation reflects cortical feedback via V1. Given its afferent input the LGN may be in an ideal strategic position to serve as an early gatekeeper in attentional gain control.

THE VISUAL SEARCH PARADIGM HAS BEEN USED TO STUDY THE ROLE OF ATTENTION IN SELECTING RELEVANT STIMULI FROM WITHIN A CLUTTERED VISUAL ENVIRONMENT

So far this chapter has considered the situation when attention is cued to a location (or a feature) in the visual field. However, one is rarely told in advance to attend to a particular location. Normally, one needs to find a particular object in a complex world that is composed of a large number of stimuli. Psychologists have used a visual search paradigm to understand how attention selects behaviorally relevant stimuli out of a group of other stimuli.

In a typical task, observers are asked to search among an array of stimuli and indicate whether a particular target is present in the array. This task is easier under some conditions than others. For example, it is easy to determine whether a horizontal green bar is present in Figure 46.11A. It takes longer to make this judgment for stimuli in Figure 46.11B. Treisman and Gelade (1980) found that searches for a target with a unique feature, like the one illustrated in Figure 46.11A, can be completed quickly regardless of the number of elements in the search array; the target seems to "pop out" of the search array. Such searches often are referred to as efficient. In contrast, Treisman and Gelade (1980) found that the amount of time required to find a target that is defined by conjunctions of elementary features (e.g., a horizontal green among horizontal reds and vertical greens, as in Fig. 46.11B) increases linearly with the number of elements in the array. Such searches often are referred to as inefficient. The amount of time added per item depends on stimulus conditions, but a rule of thumb is that each additional item in the array adds about 50 ms to the amount of time taken to locate the target. The fact that it takes time to locate the target in Figure 46.11B implies that the visual system is limited in capacity. If it were not, the target could be identified immediately by evaluating every element in the array simultaneously.

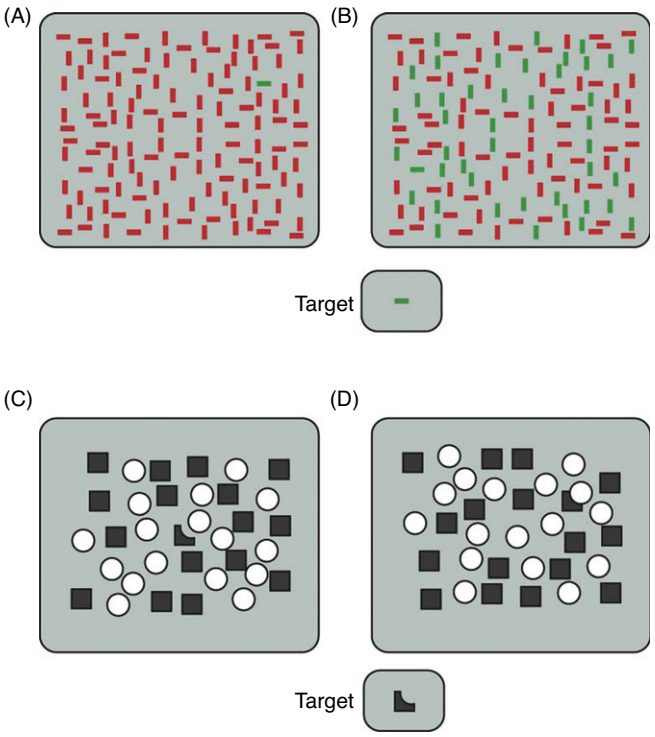


FIGURE 46.11 Efficient and inefficient visual searches. (A) Pop-out search. The horizontal green bar differs from all other array elements in a single feature, color. As a result, it pops out immediately, regardless of how many elements are in the array. (B) The same target is difficult to find when it is embedded among elements that cannot be differentiated from it on the basis of a single feature (color or orientation). (C) The dark target pops out among squares and circles. (D) When the same target is perceived to be a square occluded behind a circle, it is difficult to find among squares and circles. Because this arrangement of the array interferes with visual search, this suggests that integration of objects into wholes occurs prior to the activation of attentional mechanisms used in visual search. Thus, these mechanisms are evidently not necessary for completion. *Adapted from Rensink and Enns (1998).*

WHERE IS THE COMPUTATIONAL BOTTLENECK AS REVEALED BY SEARCH TASKS?

Considerable debate exists regarding the locus and nature of the computational bottleneck that causes the capacity limitations that are revealed by visual search tasks. One influential theory, feature integration theory (FIT; Treisman & Gelade, 1980), proposes that capacity limitations occur at very early stages of cortical processing where "elementary" features are bound into coherent objects. This theory proposes that in the "pop-out" case, where the target differs from distracters in a single elemental feature, a search can be completed efficiently because there is no need to bind together different elementary features. Thus, a green target among red distracters (Fig. 46.11A) can be found in parallel by V1 neurons that respond to green but not red stimuli, without the need

for visual attention. However, in a conjunction search, where each of the elementary target features (color and orientation) is shared with distracters, the observer must bind elementary features into coherent object representations before search can occur. According to this account, one cannot process multiple objects simultaneously because in order to avoid misconjoining features from different objects, attention selects features from one location at a time. Consistent with this idea, patient RM, who suffered damage to his parietal attentional control system, often misconjoins the shape and color of letters, even after viewing them for up to 10 seconds; that is, he mixes up which letters are presented in which colors.

However, the idea that limited capacity occurs at early stages of processing has been challenged. For example, some studies suggest that parts of objects are integrated into wholes prior to the application of attention. Consider the task of searching for the oddly shaped object appearing at the bottom of Figure 46.11 (adapted from Rensink & Enns, 1998). This odd shape pops out from among the squares and circles in Figure 46.11C, but it is harder to find the same shape in Figure 46.11D. The difference is that in Figure 46.11D, the oddly shaped part appears to be a square that is occluded by a superimposed circle; that is, the visual system completes the shape of the occluded square, hiding it among the other squares. Because this completion stops the square from popping out, the completion operation must occur prior to visual search.

Further evidence that features can be integrated preattentively into objects comes from the finding that when an observer attends to one feature of an object (e.g., its orientation), other features of the object are selected automatically as well. In a now classic study, Duncan (1984) found that observers could easily make simultaneous judgments about two features of the same object (e.g., its orientation and whether or not it contained a gap). Making the second judgment did not interfere with the first judgment. However, observers were severely impaired when they simultaneously judged the same two features, one on each of two different stimuli. This finding suggests that when attention is directed to one feature of an object, all the features that make up the object can automatically be selected together because the features are already linked together into objects. These sorts of observations suggest that limited capacity results from information processing bottlenecks that occur relatively late in processing, after features are already integrated into wholes. One possibility is that the bottleneck occurs when objects enter a limited capacity working memory. Under this view, attention plays the role of selecting which objects pass through the bottleneck.

These two perspectives are not mutually exclusive. Some feature integration may occur preattentively (e.g., Rensink & Enns, 1998), but other types of stimulus integration might require spatial selection. As described earlier, attention has been found to modulate neural activity at all stages of visual processing including early stages (Motter, 1993; O'Connor et al., 2002), where neuronal responses first show signs of grouping the features and parts of objects into wholes, as well as late stages, which are likely involved in selecting stimuli for storage in working memory. Thus, both early and late stages of processing may act as limited capacity bottlenecks.

NEURONAL RECEPTIVE FIELDS ARE A POSSIBLE NEURAL CORRELATE OF LIMITED CAPACITY

Extracellular recording studies of attention in awake, behaving monkeys have revealed that the receptive fields of individual neurons might contribute to the limitation in visual capacity. As visual information traverses the successive cortical areas of the ventral stream that underlie object recognition (Ungerleider & Mishkin, 1982), the sizes of receptive fields increase from less than a degree of visual arc in primary visual cortex (area V1) to about 20° of visual arc in area TE, the last purely visual area in the ventral visual processing stream (see Chapter 44). Similar increases in receptive field size are observed in the dorsal visual processing stream. In a typical scene, such large receptive fields will contain many different objects. Therefore, a likely explanation for why one cannot process many different objects in a scene simultaneously is that neurons, whose signals are limited in bandwidth, cannot simultaneously send signals about all the stimuli inside their receptive fields. This idea implies that processing limitations exist at all levels of processing but that they become more pronounced in higher order visual areas, where receptive field sizes are larger. How then do neurons transmit signals about behaviorally relevant stimuli in their receptive fields?

One proposal is that the selective processing of behaviorally relevant stimuli is accomplished by two interacting mechanisms: competition among potentially relevant stimuli and biases that determine the outcome of this competition. For example, when an observer is asked to detect the appearance of a target at a particular location, this task is thought to activate spatially selective feedback signals in cortical areas, such as parts of the parietal and frontal cortex discussed earlier. These cortical areas provide a task-appropriate spatial reference frame and transmit signals to the extrastriate cortex, where they bias competition in favor of stimuli that appear at the attended location. Similarly, when an observer searches for an object in a cluttered scene, feature-selective feedback is thought to bias competition in favor of stimuli that share features in common with the searched-for object. When competition is

resolved, the winning stimulus controls neuronal responses, and other stimuli are effectively filtered out of the visual stream.

Studies in monkeys have compared the effect of attention on neuronal responses with single and multiple stimuli in the receptive field. Consistent with the proposal just outlined, larger changes in firing rate occur when multiple stimuli appear within the receptive field (and one stimulus must be attended), as compared to when the attended stimulus appears alone and there is no competition to be resolved (Moran & Desimone, 1985). According to the biased competition model just described, neurons that respond to a stimulus should be suppressed when a second stimulus is added that activates a competing population of neurons. To test this idea, pairs of stimuli were presented that activated two nearby populations of neurons (Reynolds, Chelazzi, & Desimone, 1999). Monkeys attended away from the receptive fields of the neurons. The first stimulus was chosen to be of a color and orientation that would strongly activate the recorded neuron. Then, a second stimulus, which elicited a response in a nearby population of neurons, was added. As predicted, the addition of this second stimulus partially suppressed the responses of many neurons to the first stimulus. To test whether attentional feedback would resolve competition in favor of the neurons that responded to the attended stimulus, attention was then directed to the poor stimulus. As predicted, this often magnified the suppressive influence of the poor stimulus, further reducing the firing rates of many neurons (Fig. 46.12A). In a final experiment, attention was directed to the preferred stimulus. Consistent with the resolution of competition in favor of the neurons that responded to the preferred stimulus, directing attention to the preferred stimulus strongly reduced the suppressive effect of the poor stimulus (Fig. 46.12B).

These findings, and related studies conducted over the past decade, have led to increasingly refined models of attentional selection, grounding our understanding of the automatic gain control circuitry that adjusts visual sensitivity in response to variation in the strength of sensory input (Reynolds & Chelazzi, 2004). Gain control is necessitated by the enormous range of stimulus intensity, such as luminance, luminance contrast, and color contrast that occurs within the visual environment. Take, for example, the change in the number of photons that enter the eye. This varies a million fold from a moonlit night to a sunny day. Evolution has endowed us with a variety of automatic gain control mechanisms, which, together, help us to remain sensitive to subtle changes in sensory input in the face of such enormous variation in intensity. These mechanisms range from the pupillary response to gain control mechanisms within the retina, the lateral geniculate nucleus, and the visual cortices. Decades of research in anesthetized animals have elucidated these neural

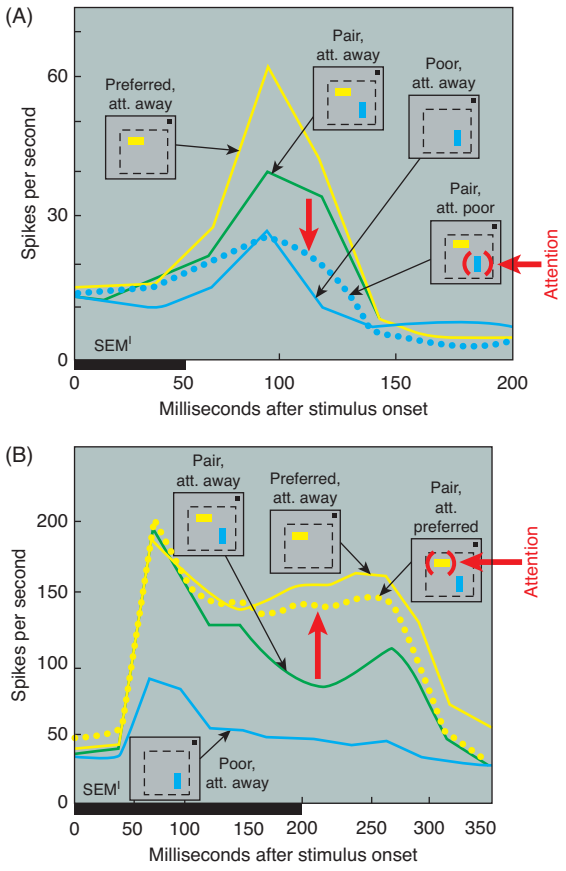


FIGURE 46.12 Attention to one stimulus of a pair filters out the effect of the ignored stimulus. (A) The x-axis shows time (in milliseconds) from stimulus onset, and the thick horizontal bar indicates stimulus duration. Small iconic figures illustrate sensory conditions. Within each icon, the dotted line indicates the receptive field, and the small dot represents the fixation point. The location of attention inside the receptive field is indicated in red. Attention was directed away from the receptive field in all but one condition. The preferred stimulus is indicated by a horizontal yellow bar and the poor stimulus by a vertical blue bar. In fact, the identity of both stimuli varied from cell to cell. The yellow line shows the response of a V2 neuron to the preferred stimulus. The solid blue line shows the response to the poor stimulus. The green line shows the response to the pair. The addition of the poor stimulus suppressed the response to the preferred stimulus. Attention to the poor stimulus (red arrow) magnified its suppressive effect and drove the response down to a level (dotted blue line) that was similar to the response elicited by the poor stimulus alone. (B) Response of a second V2 neuron. The format is the same as in A. As in the neuron above, the response to the preferred stimulus was suppressed by the addition of the poor stimulus. Attention directed to the preferred stimulus filtered out this suppression, returning the neuron to a response (dotted yellow line) similar to the response that was elicited when the preferred stimulus appeared alone inside the RF. *Adapted from Reynolds et al. (1999).*

mechanisms, and it appears that the attentional system has co-opted some of them, including suppressive circuitry within the striate and extrastriate visual cortices. Models in which these gain control circuits are modulated by feedback from attentional control centers have proven

able to account for attention-dependent changes in sensitivity, and competitive selection of task-relevant stimuli from among task-irrelevant distracters (Reynolds et al., 1999; Reynolds & Heeger, 2009). It thus appears that while automatic gain control mechanisms may originally have evolved to act in response to changes in the external environment, they also play a key role in endogenously generated attentional selection.

FILTERING OF UNWANTED INFORMATION IN HUMANS

Functional brain imaging studies suggest that in the human visual cortex, just as in the monkey cortex, mechanisms exist such that multiple stimuli compete for neural representation and unwanted information can be filtered out (Kastner, De Weerd, Desimone, & Ungerleider, 1998). In one fMRI study, subjects maintained fixation while they were presented with colorful visual stimuli in four nearby locations in the periphery of the visual field (Kastner et al., 1998). Stimuli were presented under two different conditions. In a sequential condition, a single stimulus appeared in one of the four peripheral locations, then another appeared in a different location, and so on, until each of the four stimuli had been presented in the four different locations. In a simultaneous condition, the same four stimuli appeared in the same four locations, but they were presented at the same time. Thus, physical stimuli were identical in the two conditions except that in the second case, all four stimuli appeared simultaneously, which should encourage competitive interactions among the stimuli. The subjects' task was to count letters at fixation, thereby ignoring the peripheral stimulus presentations. Activation of V1 and ventral stream extrastriate areas V2 to TEO was found under both presentation conditions. Although the fMRI signal was similar in the two presentation conditions in V1, activation was reduced in the simultaneous condition compared to the sequential condition in V2. This reduction in activity was especially pronounced in V4 and TEO. The most straightforward interpretation of this result is that simultaneously presented stimuli interacted in a mutually suppressive way. This sensory suppression effect may be a neural correlate of competition in the human visual cortex. Importantly, the suppression effects appeared to be scaled to the receptive field size of neurons within visual cortical areas; that is, the effects were larger in areas with larger receptive fields. This observation suggests that, as in the monkey visual cortex, competitive interactions occur most strongly at the level of the receptive field (Kastner et al., 1998).

The effects of spatially directed attention on multiple, competing visual stimuli have been studied in a variation of the same paradigm (Kastner et al., 1998). In addition to the two different visual presentation conditions, sequential and simultaneous, two different attentional conditions were tested, unattended and attended. During the unattended condition, attention was directed away from the visual display by having subjects count the occurrences of letters at the fixation point. In the attended condition, subjects were instructed to attend covertly to the stimulus location that was closest to the fixation point and to count the occurrences of one of the four stimuli. The attended condition led to greater increases in fMRI signals to simultaneously presented stimuli than to sequentially presented stimuli. Thus, in this case, attention partially cancelled out the suppressive interactions among competing stimuli. The magnitude of the attentional effect was greater where the suppressive interactions among stimuli had been strongest, such that the strongest reduction of suppression occurred in areas V4 and TEO. These findings support the idea that directed attention enhances information processing of stimuli at attended locations by counteracting the suppression induced by nearby stimuli that are competing for limited processing resources. In this way, unwanted distracting information is filtered out effectively.

Unwanted information can not only be filtered out effectively by top-down mechanisms that resolve competitive interactions among multiple objects, but also by bottom-up mechanisms that are defined by the context of the stimuli. As we have seen earlier, if a salient stimulus is present in a cluttered scene, it will be effortlessly and quickly detected regardless of the number of distracters, suggesting that competition is biased in favor of the salient stimulus. It has been shown that bottom-up influences related to stimulus context of a visual display such as a single, salient stimulus that pops out from a homogeneous, or several stimuli that are grouped together by a Gestalt rule such as collinearity or similarity can partially or even entirely overcome competitive interactions in a multiple-stimulus display. A pop-out display condition, in which a single item differed from the others, was compared to a heterogeneous display, in which all four stimuli were different. Subjects were engaged in a difficult letter detection task at fixation and never attended to the pop-out or heterogeneous displays that were shown in the periphery of the visual field. Competitive interactions among multiple stimuli were eliminated in area V4 when the stimuli were presented in the context of pop-out displays relative to heterogeneous displays, indicating that competition was biased in favor of the salient stimulus in a bottom-up stimulus context-dependent fashion. These pop-out effects appeared to originate from area V1, where the strongest responses to pop-out displays were found. The effects of pop-out on competitive interactions among multiple stimuli appeared to operate in an automatic fashion, thus independently of attentional top-down control. Taken together, the studies reviewed in this section

indicate that both top-down attention and bottom-up stimulus-driven mechanisms are critical in resolving competitive interactions among multiple stimuli at intermediate processing levels of visual cortex, thereby filtering out unwanted information from cluttered visual scenes.

COMPETITION CAN BE BIASED BY NONSPATIAL FEEDBACK

The findings just summarized support the proposal that competitive neural circuits can be biased by spatially selective feedback signals. As illustrated in Figure 46.13, feature-selective feedback can also bias competition. Responses of neurons in area TE were recorded as monkeys searched for a target (Chelazzi, Duncan, Miller, & Desimone, 1998). On each trial, the monkey viewed a cue stimulus that appeared at the fixation point. After a delay, an array of stimuli (the search array) appeared within the receptive fields of neurons in area TE. The monkey's task was to indicate whether a target matching the cue stimulus was present in the array, by making a saccade to the target. Locations of the stimuli were selected at random so that the monkey could not know where the target would appear, if it appeared at all. During the delay period, many TE neurons had a higher baseline firing rate when the cue stimulus was a preferred stimulus for the cell than when it was a poor stimulus; that is, neurons that were tuned to respond to the target were activated while the monkey prepared to search for the target. An interesting possibility is that this activation might bias the visual system toward detecting the target object, just as spatially specific feedback is thought to bias the system toward a particular location. Consistent with this, within 150 to 200 ms after the search array appeared, the response of the neuron increased or decreased, depending on whether the target was, respectively, a preferred or a poor stimulus for the cell. That is, shortly after the appearance of the search array, the activity of the cell reflected the identity of the target stimulus and was no longer influenced by the nontarget stimulus, even though the nontarget stimulus was still present physically within its receptive field.

Biases in the form of nonspatial feedback are potent mechanisms to overcome the clutter imposed by natural environments. For example, when we search for objects in the real world, we may form concrete (e.g., looking for a familiar face in a crowd of people) or more abstract (e.g., looking for cars when crossing the street) search templates that facilitate the detection of the behaviorally relevant information that helps to filter out the distracting clutter. Recent studies on attentional selection from natural scenes have shown that the representation of objects in visual cortex is entirely task-dependent (Peelen et al., 2009). Subjects were asked to look for people or cars in a large set of briefly presented natural scenes. Task-relevant

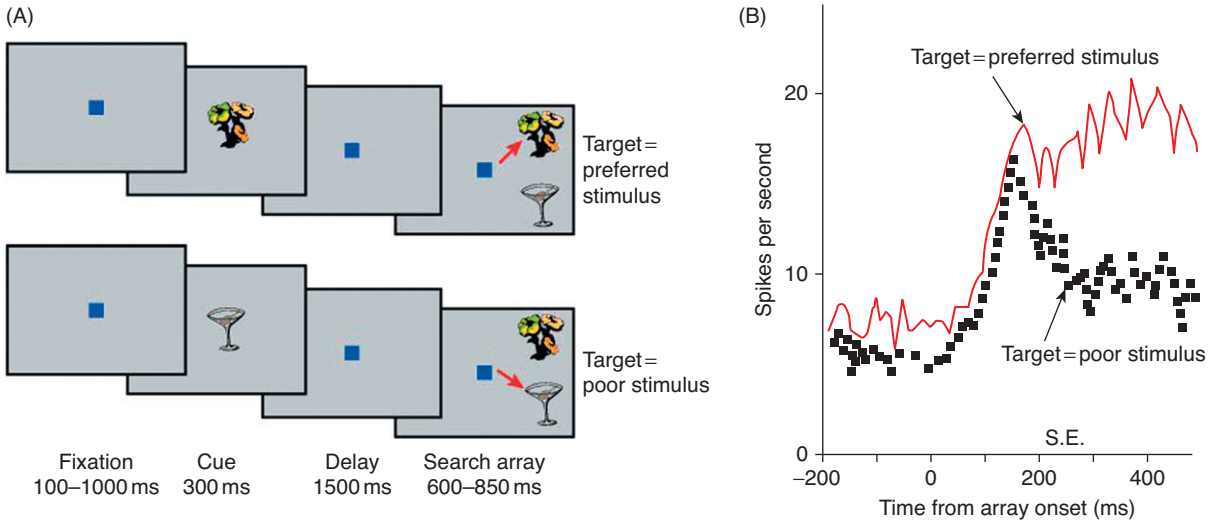


FIGURE 46.13 Responses of inferior temporal (IT) neurons during a memory-guided visual search. (A) Task. Monkeys fixated a spot on a computer screen and were shown a central cue (here, either the flower or the cup). After a delay, two (or more) stimuli appeared within the receptive field and the monkey had to saccade to the stimulus that had appeared earlier as the cue. Sometimes (top four images), the cue was a preferred stimulus for the cell (the flower) and the monkey had to saccade to the preferred stimulus. On separate trials (lower four images), the cue was a poor stimulus (the cup) and the monkey had to saccade to the poor stimulus. (B) Neuronal responses. During the delay period, IT neurons showed an elevated baseline activity that reflected the cue stored in memory. The spontaneous firing rate was higher on trials in which the cue was a preferred stimulus for the cell, relative to trials when the cue was a poor stimulus. After the search array appeared, the responses separated, increasing or decreasing depending on whether the cue was, respectively, a preferred or poor stimulus for the cell. This separation occurred well before the onset of the saccade, which is indicated by the vertical bar on the horizontal axis.

objects were represented in visual cortex, even at spatially unattended locations, while distracter objects did not get processed up to category level, even when spatially attended. Search templates for the target category were shown to pre-activate object-selective cortex in category-specific ways. The strength of preactivation could be used as a reliable predictor of task performance during the detection task. More generally, internally generated templates may provide powerful biases in generating predictions about sensory and behavioral context that enable flexible behavior.

CONCLUSIONS

The behavioral manifestations of attention are clearly not the result of operations in any one dedicated neural center. Selective attention affects neural activity at all levels of the visual system and other sensory systems, with the possible exception of neurons in the sensory periphery. A few structures located at the top of the visual system hierarchy in the parietal, frontal, and cingulate cortices are thought to be relatively more specialized for generating the signals that guide selective attention. These are among the areas that are thought to provide feedback signals that influence sensory processing in areas that are specialized for analyzing the physical properties of the sensory environment.

Answers are beginning to be available to the question of why attention is necessary in the first place. A possible computational bottleneck may be built into all levels of the visual system in the form of spatial receptive fields. The fact that several objects within a certain spatial region can activate the same neuron reduces the ability of such neurons to transmit accurate information about each one of these objects. Automatic gain control mechanisms appear to have been co-opted by the attentional system to mediate selection of attended stimuli from among task-irrelevant distracters. Other bottlenecks will doubtlessly be revealed in future research.

References

Bisley, J. W., & Goldberg, M. E. (2003, January 3). Neuronal activity in the lateral intraparietal area and spatial attention. *Science, 299* (5603), 81–86.

Bushnell, M. C., Goldberg, M. E., & Robinson, D. L. (1981). Behavioral enhancement of visual responses in monkey cerebral cortex. I. Modulation in posterior parietal cortex related to selective visual attention. *Journal of Neurophysiology, 46,* 755–772.

Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *Journal of Neurophysiology, 80*(6), 2918–2940.

Cohen, M. R., & Maunsell, J. H. R. (2009). Attention improves performance primarily by reducing interneuronal correlations. *Nature Neuroscience, 12,* 1594–1600.

Constantinidis, C., & Steinmetz, M. (2001). Neuronal responses in area 7a to multiple stimulus displays: II responses are suppressed at the cued location. *Cerebral Cortex, 11*(7), 592–597.

Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., & Petersen, S. E. (1991). Attentional modulation of neural processing of shape, color, and velocity in humans. *Science, 248,* 1556–1559.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology. General, 113*(4), 501–517.

Egeth, H. E., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology, 48,* 269–297.

Gottlieb, J. M., Kusunoki, M., & Goldberg, M. E. (1998). The representation of visual salience in monkey parietal cortex. *Nature, 391,* 481–484.

Harris, K. D., & Thiele, A. (2011). Cortical state and attention. *Nature Reviews Neuroscience, 12,* 509–523.

Heilman, K. M. (1979). Neglect and related disorders. In E. Valenstein (Ed.), *Clinical neuropsychology* (pp. 268–307). New York: Oxford University Press.

Heinze, H. J., Mangun, G. R., Burchert, W., Hinrichs, H., Scholz, M., Münte, T. F., Hillyard, S. A., et al. (1994). Combined spatial and temporal imaging of brain activity during visual selective attention in humans. *Nature, 372,* 543–546.

Hillyard, S. A., & Vento, L. A. (1998). Event-related brain potentials in the study of visual selective attention. *Proceedings of the National Academy of Sciences of the United States of America, 95,* 781–787.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews. Neuroscience, 2*(3), 194–203.

Kastner, S., & Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience, 23,* 315–342.

Kastner, S., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science, 282,* 108–111.

Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron, 22,* 751–761.

Kohn, A., & Smith, M. A. (2005). Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *Journal of Neuroscience, 25,* 3661–3673.

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research, 35*(13), 1897–1916.

Lovejoy, L. P., & Krauzlis, R. J. (2010). Inactivation of primate superior colliculus impairs covert selection for perceptual judgments. *Nature Neuroscience, 13*(2), 261–266.

McAlonan, K., Cavanaugh, J., & Wurtz, R. H. (2008). Guarding the gateway to cortex with attention in visual thalamus. *Nature, 456,* 391–394.

Mesulam, M. M. (1999). Spatial attention and neglect: Parietal, frontal and cingulate contributions to the mental representation and attentional targeting of salient extrapersonal events. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 354*(1387), 1325–1346.

Mitchell, J. F., Sundberg, K. A., & Reynolds, J. H. (2007, July 5). Differential attention-dependent response modulation across cell classes in macaque visual area V4. *Neuron, 55*(1), 131–141.

Mitchell, J. F., Sundberg, K. A., & Reynolds, J. H. (2009, September 13). Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron, 63,* 879–888.

Moore, T., & Armstrong, K. M. (2003). Selective gating of visual signals by microstimulation of frontal cortex. *Nature, 421,* 370–373.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science, 229*(4715), 782–784.

Motter, B. C. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *Journal of Neurophysiology, 70,* 909–919.

O'Connor, D. H., Fukui, M. M., Pinsk, M. A., & Kastner, S. (2002). Attention modulates responses in the human lateral geniculate nucleus. *Nature Neuroscience, 5,* 1203–1209.

Oristaglio, J., Schneider, D. M., Balan, P. F., & Gottlieb, J. (2006). Integration of visuospatial and effector information during symbolically cued limb movements in monkey lateral intraparietal area. *Journal of Neuroscience, 26*(32), 8310–8319.

Peck, C. J., Jangraw, D. C., Suzuki, M., Efem, R., & Gottlieb, J. (2009). Reward modulates attention independently of action value in posterior parietal cortex. *Journal of Neuroscience, 29*(36), 11182–11191.

Peelen, M. V., & Fei-Fei, N. P. (2009). Neural mechanisms of rapid scene categorization in human visual cortex. *Nature, 460,* 94–98.

Posner, M. I., Walker, J. A., Friedrich, F. J., & Rafal, R. D. (1984). Effects of parietal lobe injury on covert orienting of visual attention. *Journal of Neuroscience, 4,* 1863–1874.

Rensink, R. A., & Enns, J. T. (1998). Early completion of occluded objects. *Vision Research, 38*(15–16), 2489–2505.

Reynolds, J. H., & Chelazzi, L. (2004). Attentional modulation of visual processing. *Annual Review of Neuroscience, 27,* 611–647.

Reynolds, J. H., & Heeger, D. J. (2009, Jan 29). The normalization model of attention. *Neuron, 61*(2). 168–85.

Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *Journal of Neuroscience, 19*(5), 1736–1753.

Sugrue, L. P., Corrado, G. S., & Newsome, W. T. (2004, June 18). Matching behavior and the representation of value in the parietal cortex. *Science, 304*(5678), 1782–1787.

Thompson, K. G., & Bichot, N. P. (2005). A visual salience map in the frontal eye field. *Progress in Brain Research, 147,* 251–262.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97–136.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, R. J. W. Mansfield & M. A. Goodale (Eds.), *The analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.

Vallar, G. (1993). The anatomical basis of spatial neglect in humans. In I. H. Robertson & J. C. Marshall (Eds.), *Unilateral neglect: Clinical and experimental studies* (pp. 27–62). Hillsdale, NJ: Lawrence Erlbaum.

Yantis, S., & Serences, J. (2003). Neural mechanisms of space-based and object based attentional control. *Current Opinion in Neurobiology, 13,* 187–193.

Zohary, E., Shadlen, M. N., & Newsome, W. T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature, 370,* 140–143.

Suggested Readings

Connor, C. E., Preddie, D. C., Gallant, J. L., & Van Essen, D. C. (1997). Spatial attention effects in macaque area V4. *Journal of Neuroscience, 17*(9), 3201–3214.

Corbetta, M. (1998). Frontoparietal cortical networks for directing attention and the eye to visual locations: Identical, independent, or overlapping neural systems? *Proceedings of the National Academy of Sciences of the United States of America, 95*(3), 831–838.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18,* 193–222.

Driver, J., Davis, G., Russell, C., Turatto, M., & Freeman, E. (2001). Segmentation, attention and phenomenal visual objects. *Cognition, 80,* 61–95.

Heeger, D. J. (1999). Linking visual perception with human brain activity. *Current Opinion in Neurobiology, 9*(4), 474–479.

Knudsen, E. I. (2007). Fundamental components of attention. *Annual Review of Neuroscience, 30,* 57–78.

Posner, M. I. (1994). Attention: The mechanisms of consciousness. *Proceedings of the National Academy of Sciences of the United States of America, 91*(16), 7398–7403.

Treue, S. (2001). Neural correlates of attention in primate visual cortex. *Trends in Neurosciences, 24*(5), 295–300.

John H. Reynolds, Jacqueline P. Gottlieb, and Sabine Kastner

This page intentionally left blank

CHAPTER

47

Learning and Memory: Basic Mechanisms

Significant advances have been made in understanding the ways in which the nervous system encodes and retrieves information. One focus of current research is on the understanding of learning and memory at the cellular level, where the information-encoding process can be traced to changes in neuronal properties such as membrane excitability and synaptic strength. One emerging principle is that no single universal mechanism for learning and memory exists. Instead, different memory systems (see Chapter 48) can use different mechanisms. Therefore, a comprehensive understanding of memory mechanisms requires an understanding of the general ways in which neurons are changed by learning and the ways in which those changes are maintained and expressed at the cellular level.

Research during the past several decades on several vertebrate and invertebrate model systems has led to the development of several general principles. A list of these principles (Byrne, 1987) might include the following:

1. Multiple memory systems are present in the brain (see also Chapter 48).
2. Short-term and long-term forms of learning and memory involve changes in existing neural circuits.
3. These changes may involve multiple cellular mechanisms within individual neurons.
4. Second-messenger systems play a role in mediating cellular changes.
5. Changes in membrane channels are often correlated with learning and memory.
6. Long-term memory requires new protein synthesis, whereas short-term memory does not.

This chapter describes several types of neural and molecular mechanisms implicated in learning and memory. The chapter first considers some of the paradigms that have been used to study simple forms of nonassociative and associative learning and provides an example of mechanistic analyses that have been performed in a selected invertebrate model system. The later sections of the chapter describe the mechanisms of two phenomena, which are known as long-term potentiation (LTP) and long-term depression (LTD). Both LTP and LTD occur in forebrain structures, and LTP and LTD are thought to be mechanisms for memory storage in the central nervous system.

PARADIGMS HAVE BEEN DEVELOPED TO STUDY ASSOCIATIVE AND NONASSOCIATIVE LEARNING

Associative Learning

Associative learning is a broad category that includes many of our daily learning activities that involve the formation of associations among stimuli and/or responses. It is usually subdivided into classical conditioning and instrumental conditioning. Classical (or Pavlovian) conditioning is induced by a procedure in which a generally neutral stimulus, termed a *conditioned stimulus* (CS), is paired with a stimulus that generally elicits a response, termed an *unconditioned stimulus* (US). Two examples of unconditioned stimuli are food, which elicits salivation, or a shock to the foot, which elicits limb withdrawal. Instrumental (or operant) conditioning is a process by which an organism learns to associate consequences with its own behavior. The delivery of a reinforcing stimulus is contingent upon the expression of a designated behavior. The probability that this behavior will be expressed

© 2013 Elsevier Inc. All rights reserved.

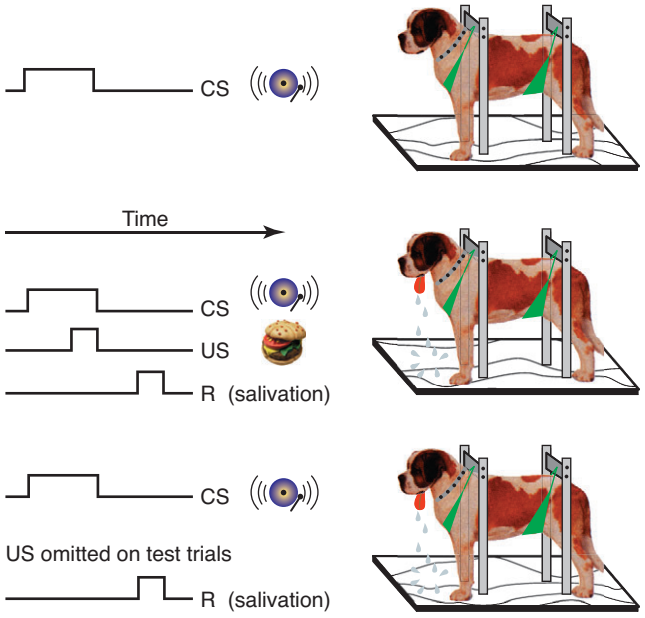


FIGURE 47.1 Classical conditioning. In the procedure introduced by Pavlov, the production of saliva is monitored continuously. Presentation of meat powder (US) reliably leads to salivation, whereas some "neutral" stimulus such as a bell (CS) initially does not. With repeated pairings of the bell and meat powder, the animal learns that the bell predicts the food and salivates in response to the bell alone. *Modified from Rachlin (1991).*

is then altered. This chapter focuses on classical conditioning, as it is mechanistically the best understood type of associative conditioning.

An astute observation by Ivan Pavlov, a Russian physiologist who had been studying digestion in dogs, led to his discovery of classical conditioning. He first noticed that the mere sight of the food dish caused dogs to salivate. He continued the experiments to see if dogs would also salivate in response to a bell rung at feeding time. Pavlov trained dogs to stand in a harness and, after the sound of a bell, fed them meat powder (Fig. 47.1). At first, the bell by itself did not elicit any response, but the meat powder elicited reflex salivation, which was termed the *unconditioned response* (UR). He noted that after a few pairings of the bell and meat powder the dogs began to salivate when the bell rang, before they received the meat powder. This response is termed the *conditioned response* (CR). This type of conditioning came to be called reward or appetitive *classical conditioning*. If the bell or another stimulus was followed by an unpleasant event, such as an electric shock, then a variety of autonomic responses became conditioned. This type of conditioning is termed aversive conditioning.

Nonassociative Learning

Three examples of nonassociative learning have received the most experimental attention: habituation, dishabituation, and sensitization. *Habituation* is defined as a reduction in the response to a stimulus that is delivered repeatedly. *Dishabituation* refers to the restoration or recovery of a habituated response due to the presentation of another, typically strong, stimulus to the animal. *Sensitization* is an enhancement or augmentation of a response produced by the presentation of a strong stimulus. The following sections focus on the neural and molecular mechanisms of sensitization.

INVERTEBRATE STUDIES: KEY INSIGHTS FROM *APLYSIA* INTO BASIC MECHANISMS OF LEARNING

Since the mid-1960s, the marine mollusc *Aplysia* has proven to be an extremely useful model system to gain insights into the neural and molecular mechanisms of simple forms of memory. Indeed, the pioneering discoveries of Eric Kandel using this animal were recognized by his receipt of the Nobel Prize in Physiology or Medicine in 2000. A number of characteristics make *Aplysia* well suited for the examination of the molecular, cellular, morphological, and network mechanisms underlying neuronal modifications (plasticity) and learning and memory. The animal has a relatively simple nervous system with large, individually identifiable neurons that are accessible for detailed analyses. Indeed, the neurons and neural circuits that mediate many behaviors in *Aplysia* have been identified. In several cases, these behaviors have been shown to be modifiable by learning. Moreover, specific loci within neural circuits at which modifications occur during learning have been identified, and aspects of the cellular mechanisms underlying these modifications have been analyzed (For reviews see Bailey & Kandel, 2008; Byrne & Kandel, 1996; Fioravante, Antzoulatos, & Byrne, 2008; Kandel, 2001).

Sensitization of withdrawal reflexes has been particularly well studied. A single sensitizing stimulus, such as a brief several second-duration electric shock, can produce a reflex enhancement that lasts minutes (short-term sensitization), whereas prolonged training (e.g., multiple stimuli over an hour or more) produces an enhancement that lasts from days to weeks (long-term sensitization).

Multiple Cellular Processes Mediate Short- and Long-Term Sensitization in *Aplysia*

A discussion of memory mechanisms can be divided into three parts: induction, expression, and maintenance. *Induction* refers to the initial events that trigger or initiate the modification process; *expression* concerns how the modification process is ultimately expressed; and *maintenance* addresses the manner in which the modification is made to endure over time.

Short-Term Sensitization

Short-term (minutes) sensitization is induced when a single, brief train of shocks is delivered to the body wall of the animal. This process results in the release of modulatory transmitters, such as serotonin (5-HT), from a distinct class of interneurons (IN) referred to as facilitatory neurons (Fig. 47.2A1). These facilitatory neurons regulate the properties of the sensory neurons (SN) and the strength of their connections with postsynaptic interneurons and motor neurons (MN) through a process called *heterosynaptic facilitation* (Byrne & Kandel, 1996; Figs. 47.2A2 and 47.2A3). The molecular mechanisms contributing to short-term heterosynaptic facilitation are illustrated in Figure 47.2B. Serotonin binds to multiple types of G-protein-coupled receptors localized in the membrane of SNs and activation of these receptors recruits at least two predominant signaling cascades in the SNs: the cyclic adenosine monophosphate/protein kinase A (cAMP/PKA) and the diacylglycerol/protein kinase C (DAG/PKC) cascades (Byrne & Kandel, 1996). Calcium/calmodulin-dependent protein kinase (CaM kinase II) appears to be engaged as well. PKA and PKC have multiple targets, including membrane ion channels, thus affecting the spike waveform and the excitability of the SN. For example, activation of the PKA and PKC cascades modifies various K^+ channels, leading to an enhanced spike activity (i.e., enhanced excitability), and broadening of the action potential. Reflections of enhanced excitability include an increase in the number of action potentials elicited in a sensory neuron by a fixed extrinsic current injected into the cell or by a fixed stimulus to the skin (Fig. 47.2A3). Spike broadening leads to enhanced Ca^{2+} influx in presynaptic terminals and facilitates transmitter release. In addition to spike broadening, spike-duration-independent mechanisms of enhanced transmitter release have been implicated in 5-HT-induced heterosynaptic facilitation. These mechanisms involve direct modulation of a presynaptic calcium conductance independently of spike broadening (Leal & Klein, 2009) as well as an increase in the number of transmitter-containing synaptic vesicles that become available for release, possibly through synapsin-regulated vesicle mobilization (Fig. 47.2; Byrne & Kandel, 1996; Fioravante, Liu, Netek, Cleary, & Byrne, 2007; Zhao & Klein, 2002). This process is represented in Figure 47.2B (large open arrow) as the translocation or mobilization of transmitter vesicles from a reserve pool to a releasable pool results in an increase in the number of transmitter-containing vesicles available for release, with subsequent action potentials in the sensory neuron. The maintenance of short-term sensitization is dependent on the persistence of the PKA- and PKC-induced phosphorylations of the various substrate proteins. Prolonged treatments of 5-HT (1.5 h) also activate mitogen-activated protein kinase (MAPK) (for review, see Sharma & Carew, 2004). However, this pathway appears to be important only for the induction of longer-term processes (see section on Long-term Sensitization).

The consequences of activating these multiple second-messenger systems and modulating these various cellular processes are expressed when test stimuli elicit action potentials in the sensory neuron at various times after the presentation of the sensitizing stimuli (Fig. 47.2A3).

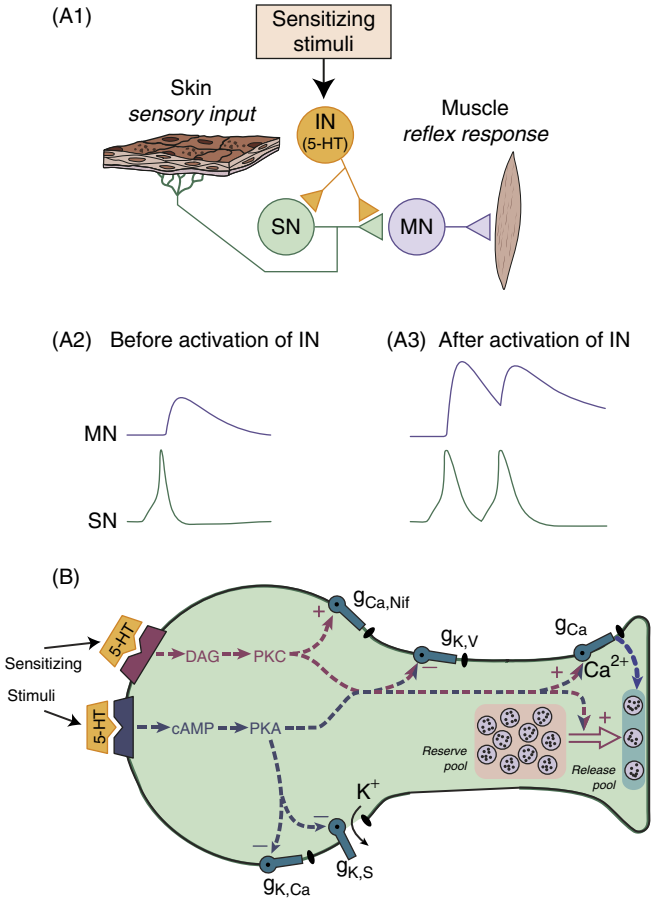


FIGURE 47.2 Model of short-term heterosynaptic facilitation of the sensorimotor connection that contributes to short-term sensitization in *Aplysia*. (A1) Sensitizing stimuli activate facilitatory interneurons (IN) that release modulatory transmitters, one of which is 5-HT. The modulator leads to an alteration of the properties of the sensory neuron (SN). (A2 and A3) The enhanced synaptic input to the motor neuron (MN) (A3) results from enhanced sensory input, partly due to two mechanisms. First, the same peripheral stimulus can evoke a greater number of action potentials in the presynaptic SN (i.e., enhanced excitability). Second, each action potential fired by an SN produces a stronger synaptic response in the MN (i.e., synaptic facilitation). (B) Model of a SN that depicts the multiple processes for short-term facilitation and changes in excitability that contribute to short-term sensitization. 5-HT released from facilitatory neurons binds to at least two distinct classes of receptors on the outer surface of the membrane and leads to the transient activation of two intracellular second messengers, DAG and cAMP, and their respective kinases (PKC and PKA). These kinases affect multiple cellular processes, the combined effects of which lead to changes in excitability and enhanced transmitter release when subsequent action potentials are fired in the sensory neuron (see text for additional details). *Modified from Byrne and Kandel (1996).*

More transmitter is available for release as a result of the mobilization process and each action potential results in a larger influx of Ca^{2+} to trigger release of the available transmitter. This facilitation of transmitter release from the sensory neuron leads to a larger postsynaptic potential in the motor neuron. Larger postsynaptic potentials lead to enhanced activation of interneurons and motor neurons and thus to an enhanced behavioral response.

Of general significance is the observation that a single modulatory transmitter (i.e., 5-HT) activates multiple kinase systems. The involvement of multiple second messenger systems in synaptic plasticity also appears to be a theme emerging from mammalian studies. For example, as discussed in a later section of the chapter, the induction of LTP in the CA1 area of the hippocampus appears to involve MAPK, PKC, CaM kinase II , and tyrosine kinase (reviewed in Dineley et al., 2001; Miyamoto, 2006).

Long-Term Sensitization

Sensitization also exists in a long-term form, which persists for at least 24 h. Whereas short-term sensitization can be produced by a single brief stimulus, the induction of long-term sensitization requires a more extensive training period over an hour or more.

Both short- and long-term sensitization share some common cellular pathways during their *induction.* For example, both forms activate the cAMP/PKA cascade (Fig. 47.3). However, in the long-term form, unlike the short-term form, activation of the cAMP/PKA cascade induces gene transcription and new protein synthesis (Kandel, 2001; Fioravante et al., 2008). Repeated training leads to a translocation of PKA to the nucleus, where it phosphorylates the transcriptional activator CREB1 (cAMP responsive element binding protein1). CREB1 binds to a regulatory region of genes known as a CRE (cAMP responsive element). Whereas CREB1 acts as an *activator* of gene transcription, a related transcription factor, CREB2, acts constitutively as a *repressor* of gene transcription. This repression of transcription is in turn *suppressed* by MAPK phosphorylation of CREB2 (Guan et al., 2002). MAPK is activated by multiple pathways including a 5-HT sensitive tyrosine receptor kinase-like molecule (ApTrk), and through feedback pathways involving the peptide sensorin and transforming growth factor β (TGF-β) (see following).

The role of transcription factors in long-term memory formation is not limited to the induction phase but may also extend to the consolidation phase, where consolidation is defined as the time window during which RNA and protein synthesis are required for converting short- to long-term memory. For example, treatment of ganglia with 5 pulses of 5-HT over a 1.5 h period to mimic sensitization training leads to the binding of CREB1 to the promoter of its own gene and induces CREB1 synthesis. The necessity of prolonged transcription and translation for long-term facilitation (LTF) observed 24 h after treatment persists for at least 10 h after induction. These results suggest that CREB1 can regulate its own level of expression, giving rise to a CREB1 positive feedback loop that supports memory consolidation (Liu, Cleary, & Byrne, 2011).

The combined effects of activation of CREB1 and removal of CREB2's repression of transcription lead to changes in the synthesis of specific proteins. So far, more

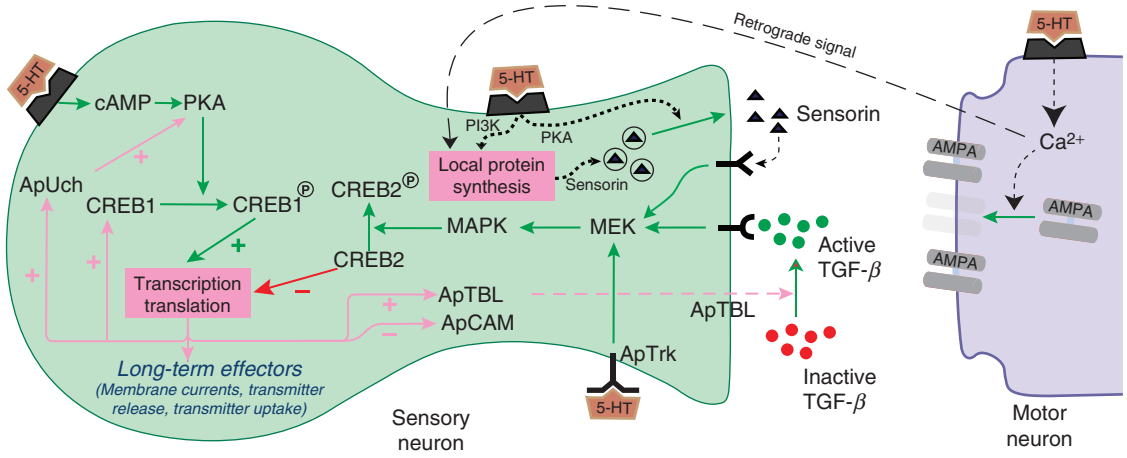


FIGURE 47.3 Simplified scheme of the mechanisms in SNs that contribute to long-term sensitization. Sensitization training leads to cAMP-dependent regulation of CREB1. Serotonin also leads to activation of MAPK, which regulates CREB2. Whereas CREB1 acts as an initiator of gene transcription, CREB2 acts as a repressor of gene transcription. The combined effects of activation of CREB1 and suppression of CREB2 lead to regulation of the synthesis of at least 10 proteins, only some of which are shown. ApTBL is believed to activate latent forms of TGF-β, which can then bind to receptors on the SN. TGF-β activates MAPK, which may act by initiating a second round of gene regulation by affecting CREB2-dependent pathways. Serotonin can also increase the release of the peptide sensorin, which binds to autoreceptors leading to further activation of MAPK. Increased local synthesis and subsequent release of sensorin can also be evoked by 5-HT, through the PI3 kinase and type II PKA, respectively. Because increased synthesis of sensorin requires elevation of postsynaptic calcium, a retrograde signal is also postulated. In addition to the retrograde signal, 5-HT-induced postsynaptic signaling also leads to an increased number of glutamate receptors.

than 10 gene products that are regulated by sensitization training have been identified, and others are likely to be found in the future. These results indicate that there is not a single memory gene or protein, but that multiple genes are regulated, and they act in a coordinated way to alter neuronal properties and synaptic strength. The following section discusses several regulated proteins of particular significance.

The downregulation of a homologue of a neuronal cell adhesion molecule (NCAM), denoted ApCAM, plays a key role in long-term facilitation. This downregulation has two components. First, the synthesis of ApCAM is reduced. Second, preexisting ApCAM is internalized via increased endocytosis. The internalization and degradation of ApCAM allow for the restructuring of the axon arbor (Bailey & Kandel, 2008). The SN can now form additional connections with the same postsynaptic target or make new connections with other cells. In addition, the presynaptic cell-adhesion protein neurexin, along with its postsynaptic counterpart neuroligin and their transsynaptic interaction, are required for 5-HT-induced LTF and the associated presynaptic structural changes (Choi et al., 2011). Furthermore, the stabilization of new structures depends on a translation-regulating protein, the *Aplysia* homolog of cytoplasmic polyadenylation element-binding protein (ApCPEB; Miniaci et al., 2008). ApCPEB appears to have prion-like properties in that it can appear in one of two or more conformations, one of which dominates and allows ApCPEB to self-perpetuate. Conversion of ApCPEB to the self-perpetuating state is enhanced by 5-HT and required for the persistence of LTF (Si, Kandel, Majumdar, Grindley, & Choi, 2010). Finally, it has been suggested that persistence of long-term facilitation and sensitization can be disrupted by inhibition of an *Aplysia* homolog of the protein kinase M (PKM Apl III; Cai, Pearce, Chen, & Glanzman, 2011). PKM (PKMζ) is a constitutively active PKC isoform that phosphorylates vertebrate AMPA-type glutamate receptors and enhances their responses, thus supporting the persistence of long-term potentiation of synapses (Sacktor, 2011).

Another protein whose synthesis is regulated by long-term facilitation is *Aplysia* tolloid/BMP-like protein (ApTBL-1). Tolloid and the related molecule BMP-1 appear to function as secreted Zn^{2+} proteases. In some systems, they activate members of the transforming growth factor β (TGF-β) family. Indeed, in SNs, TGF-β mimics the effects of 5-HT in that it produces long-term increases in the synaptic strength of the SNs (Zhang, Endo, Cleary, Eskin, & Byrne, 1997). Interestingly, TGF-β activates MAPK in the SNSs and induces its translocation to the nucleus. Thus, TGF-β could be part of an *extracellular* positive feedback loop, possibly leading to another round of protein synthesis (Fig. 47.3) to further consolidate the memory (Zhang, Endo, Cleary, Eskin, & Byrne, 1997). Another extracellular positive feedback loop involves the 5-HT-induced regulation of the release of the SN-specific neuropeptide sensorin (Fig. 47.3). Synthesis of sensorin is stimulated by 5-HT in a PI3 kinase (Phosphatidylinositol 3-kinase)-dependent manner and somewhat surprisingly, requires elevation of postsynaptic calcium (Cai, Chen, & Glanzman, 2008; Hu, Wu, & Schacher, 2006). The mechanism through which postsynaptic calcium regulates presynaptic local protein synthesis of sensorin remains unclear, but the release of a retrograde signal has been postulated (Cai et al., 2008). Interestingly, although the existence of a postsynaptic neuron seems to be required for long-term facilitation, it is not required for another correlate of long-term sensitization, increased SN excitability (Cleary, Lee, & Byrne, 1998; Liu et al., 2011). Sensorin binding to presynaptic autoreceptors activates MAPK (Hu et al., 2006), which phosphorylates CREB2 and contributes to transcriptional regulation during the consolidation of LTF.

Another important protein, *Aplysia* ubiquitin hydrolase (ApUch), appears to be involved in an *intracellular* positive feedback loop. During the induction of long-term facilitation, ApUch levels in SNs are increased, possibly via CREB phosphorylation and a consequent increase in *ApUch* transcription. The increased levels of ApUch increase the rate of degradation of proteins, via the ubiquitin-proteosome pathway, including the regulatory subunit of PKA (Chain et al., 1999). The catalytic subunit of PKA, when freed from the regulatory subunit, is highly active. Thus, increased ApUch will lead to an increase in PKA activity and a more protracted phosphorylation of CREB1. This phosphorylated CREB1 may act to further prolong *ApUch* expression, thus closing a positive feedback loop. Protein degradation, in general, and the role of ubiquitination in particular, is an emerging theme in recent studies on the neural basis of long-term memory (Fioravante & Byrne, 2011).

One simplifying hypothesis is that the mechanisms underlying the *expression* of short- and long-term sensitization are the same, but extended in time for long-term sensitization. Some evidence supports this hypothesis. For example, long-term sensitization, like short-term sensitization, is associated with an enhancement of sensorimotor connections. In addition, K^+ currents and the excitability of sensory neurons are modified by long-term sensitization (Cleary et al., 1998). However, the mechanisms underlying the *expression* of short- and long-term sensitization differ in four fundamental ways. First, long-term sensitization training leads to a decrease in the duration of the SNs action potential (Antzoulatos & Byrne, 2007). Second, structural changes such as neurite outgrowth and active zone remodeling are associated with the expression of long-term but not with short-term sensitization (Bailey & Kandel, 2008). Interestingly, the structural changes take time to develop. Long-term synaptic facilitation

measured one day after four days of training is associated with structural changes, whereas long-term facilitation measured one day after a single training session is not (Wainwright, Byrne, & Cleary, 2004). Third, long-term sensitization is associated with an increase in high-affinity glutamate uptake (Levenson et al., 2000). A change in glutamate uptake could potentially exert a significant effect on synaptic efficacy by regulating the amount of transmitter available for release, the rate of clearance from the cleft, and thereby the duration of the excitatory postsynaptic potential (EPSP) and the degree of receptor desensitization. Finally, long-term sensitization has been correlated with changes in the postsynaptic cell (i.e., the motor neuron; Cleary et al., 1998). Thus, as with other examples of memory, multiple sites of plasticity exist even within this simple reflex system.

Other Temporal Domains for the Memory of Sensitization

Historically, memory has been divided into two temporal domains, short term and long term. It has become increasing clear from studies of a number of memory systems that this distinction is overly simplistic. For example, in *Aplysia*, Carew and colleagues (Sutton, Masters, Bagnall, & Carew, 2001) and Kandel and colleagues (Ghirardi, Montarolo, & Kandel, 1995) have discovered an intermediate phase of memory that has distinctive temporal characteristics and a unique molecular signature. The intermediate-phase memory for sensitization is expressed at times approximately 5 min to 3 h after the beginning of training. Like long-term sensitization, its induction requires protein synthesis. Intermediate-term sensitization relies on both presynaptic and postsynaptic translation (protein synthesis; Antonov, Kandel, & Hawkins, 2010; Jin, Kandel, & Hawkins, 2011; Villareal, Li, Cai, & Glanzman, 2007), but not on transcription (mRNA synthesis). Intermediate-term synaptic facilitation, induced by 5-HT, requires activity of presynaptic PKC and postsynaptic CaM kinase II (Jin et al., 2011). In contrast to the well-established role of presynaptic PKA in short-term facilitation, its role in intermediate-term facilitation appears to be more variable, depending on the training protocol (Antonov et al., 2010; Ghirardi et al., 1995; Jin et al., 2011). Finally, intermediate-term facilitation has been correlated with activation of previously "silent" release sites by the recruitment of synaptic vesicles to preexisting varicosities (Kim et al., 2003).

In addition to intermediate-term memory, it is likely that *Aplysia* has different phases of long-term memory. For example, increased synthesis of a number of proteins occurs at 24 h after sensitization training and some of these proteins are different from those whose synthesis is increased during and immediately after training. These results suggest that the memory for sensitization that persists for times greater than 24 h may be dependent on the synthesis of proteins occurring at 24 h and may have a different molecular signature than the 24-h memory.

Mechanisms Underlying Associative Learning of Withdrawal Reflexes in *Aplysia*

In addition to sensitization, the withdrawal reflexes of *Aplysia* are subject to classical conditioning. The short-term classical conditioning observed at the behavioral level reflects, at least in part, a cellular mechanism called *activity-dependent neuromodulation* (Hawkins, Abrams, Carew, & Kandel, 1983; Walters & Byrne, 1983). A diagram of the general scheme is presented in Figure 47.4. The US pathway is activated by a shock to the animal, which elicits a withdrawal response (the UR). When a CS is paired consistently with the US, the animal will develop a withdrawal response (CR) to the CS. Activity-dependent neuromodulation is proposed as the mechanism for this pairing-specific effect. The US activates both a motor neuron (UR) and a modulatory system. The modulatory system delivers 5-HT to all the SNs (parts of the various CS pathways), which leads to a nonspecific enhancement of transmitter release from the SNs. This nonspecific enhancement contributes to short-term sensitization (see prior discussion). SNs whose activity is temporally contiguous with the US-mediated reinforcement are additionally modulated. Spiking in a SN during the presence of 5-HT leads to changes in that cell relative to other SNs whose activity was not paired with the US. Thus, a subsequent CS will lead to an enhanced activation of the reflex (Fig. 47.4B). Figure 47.5 illustrates a more detailed model of the proposed cellular mechanisms responsible for this example of classical conditioning. The modulator (US) acts by increasing the activity of adenylyl cyclase (AC), which in turn increases the levels of cAMP. Paired spiking in the SNs (CS) leads to a simultaneous increase in the level of intracellular calcium, which acting through calmodulin (CaM) greatly enhances the action of the modulator to increase AC activity and cAMP production. Thus, this system determines CS–US contiguity by a method of coincidence detection at the presynaptic terminal.

Now, consider the postsynaptic side of the synapse. The postsynaptic region contains NMDA-type receptors (see Chapters 8 and 10). These receptors need concurrent delivery of glutamate and depolarization in order to allow calcium to enter. The glutamate is provided by the activated SN (CS), and the depolarization is provided by the US (Antonov, Antonova, Kandel, & Hawkins, 2003; Roberts & Glanzman, 2003). Thus, the postsynaptic neuron provides another example of coincidence detection. The increase in intracellular calcium putatively causes a retrograde signal to be released from the postsynaptic cell to the presynaptic terminal, ultimately acting to further enhance the cAMP cascade in the SN. The overall amplification of the cAMP cascade acts to raise the level of PKA,

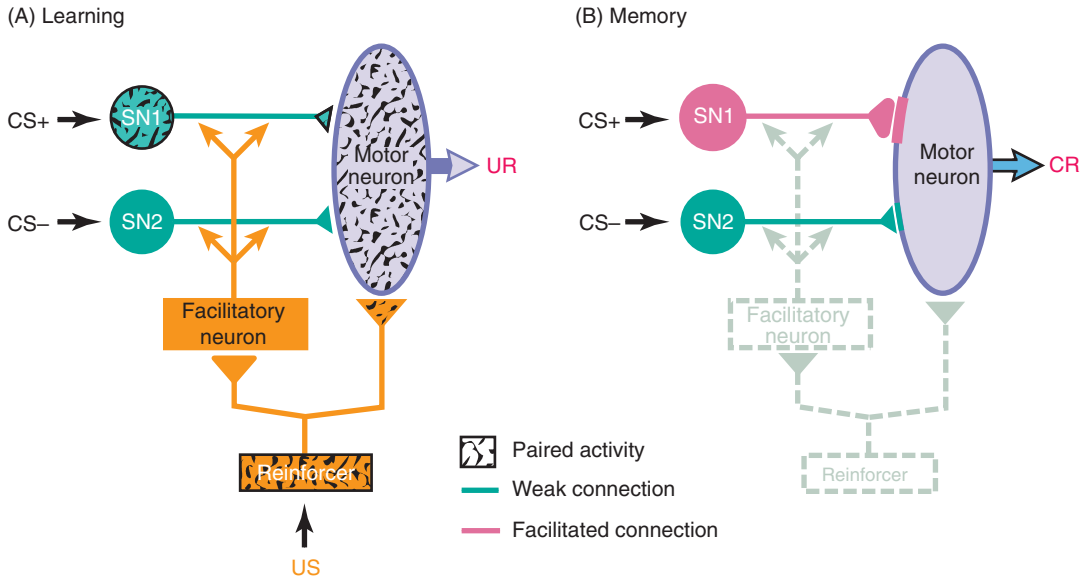


FIGURE 47.4 Model of classical conditioning of a withdrawal reflex in *Aplysia*. (A) Activity in a sensory neuron (SN1) along the CS+ (paired) pathway is coincident with activity in neurons along the reinforcement pathway (US). However, activity in the sensory neuron (SN2) along the CS– (unpaired) pathway is not coincident with activity in neurons along the US pathway. The US directly activates the motor neuron, producing the UR. The US also activates a modulatory system in the form of the facilitatory neuron, resulting in the delivery of a neuromodulatory transmitter to the two sensory neurons. The pairing of activity in SN1 with the delivery of the neuromodulator yields the associative modifications. (B) After the paired activity in A, the synapse from SN1 to the motor neuron is selectively enhanced. Thus, it is more likely to activate the motor neuron and produce the conditioned response (CR) in the absence of US input. *Modified from Lechner and Byrne (1998).*

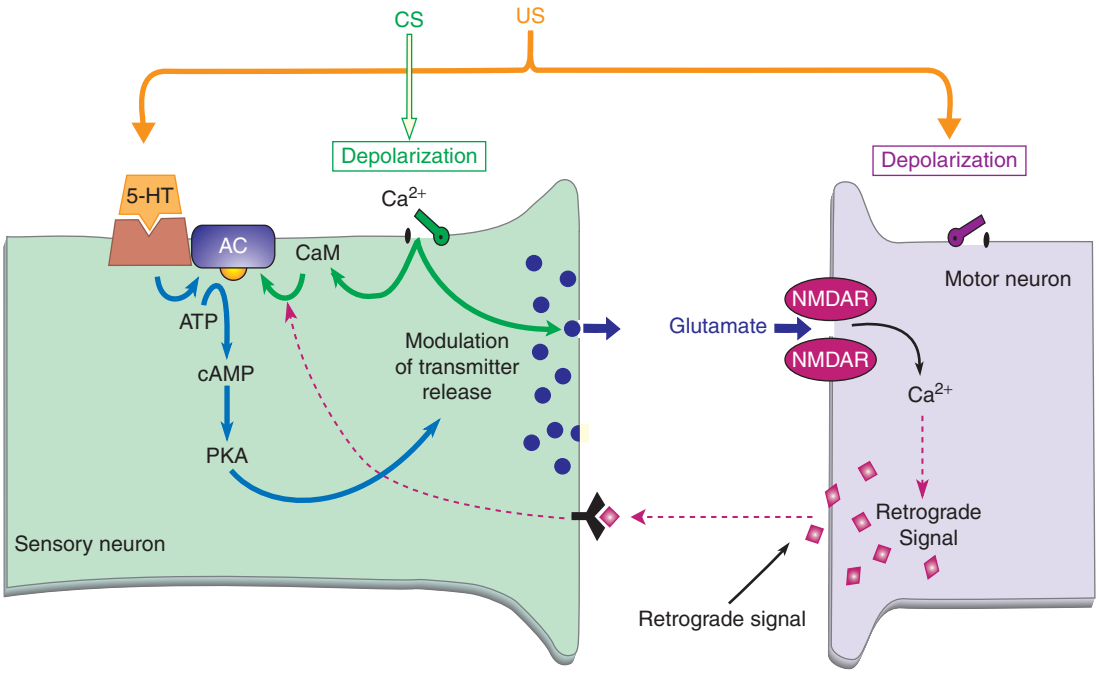


FIGURE 47.5 Model of associative facilitation at the *Aplysia* sensorimotor synapse. This model has both a presynaptic and a postsynaptic detector for the coincidence of the CS and the US. Furthermore, a putative retrograde signal allows for the integration of these two detection systems at the presynaptic level. The CS leads to activity in the sensory neuron, yielding presynaptic calcium influx, which enhances the US-induced cAMP cascade. The CS also induces glutamate release, which results in postsynaptic calcium influx through NMDA receptors if the glutamate release is paired with the US-induced depolarization of the postsynaptic neuron. The postsynaptic calcium influx putatively induces a retrograde signal, which further enhances the presynaptic cAMP cascade. The end result of the cAMP cascade is to modulate transmitter release and enhance the strength of the synapse. *Modified from Lechner and Byrne (1998).*

which in turn leads to the modulation of transmitter release. These activity-dependent changes enhance synaptic efficacy between the specific SN of the CS pathway and the motor neuron. Thus, the SN along the CS pathway will be better able to activate the motor neuron and produce the CR.

Summary

Certain invertebrates display an enormous capacity for learning and offer particular experimental advantages for analyzing the cellular and molecular mechanisms of learning. For example, behaviors in *Aplysia* are mediated by relatively simple neural circuits, which can be analyzed with conventional anatomical and electrophysiological approaches. Once the circuit is specified, the neural locus for the particular example of learning can be found, and biophysical, biochemical, and molecular approaches can then be used to identify mechanisms underlying the change. The relatively large size of some of these cells allows these analyses to take place at the level of individually identified neurons. Individual neurons can be removed surgically and assayed for changes in the levels of second messengers, protein phosphorylation, RNA, and protein syntheses. Moreover, peptides and nucleotides can be injected into individual neurons. This chapter has focused exclusively on *Aplysia,* but many other invertebrates have proven to be valuable model systems for the cellular and molecular analysis of learning and memory. Each has its own unique advantages.

Other invertebrate model systems such as *Drosophila,* although not ideal for cell biological approaches because of their small neurons, offer tremendous advantages for obtaining insights into mechanisms of learning and memory through the application of genetic approaches. A frequently used protocol in *Drosophila* employs a two-stage differential odor–shock avoidance procedure, in which animals learn to avoid odors paired (CS+) with shock but not odors explicitly unpaired (CS–). This learning is typically retained for 4–6 h, but retention for 24 h to 1 week can be produced by a spaced training procedure. Several mutants deficient in learning have been identified. Analysis of the affected genes has revealed elements of the cAMP signaling pathway as key in learning and memory. It is now known that in *Drosophila* as in *Aplysia*, the formation of long-term memory requires activation of the cAMP signaling pathway, which, in turn, activates members of the CREB transcription family. These proteins appear to be the key molecular switch that enables expression of genes necessary for the formation of long-term memories. These transcription factors are also important for long-term memory in vertebrates (see following). See Byrne (1987) and Carew (2000) for a review of other selected invertebrate model systems that have contributed importantly to the understanding of memory mechanisms.

VERTEBRATE STUDIES: LONG-TERM POTENTIATION

In contrast to the studies on invertebrates like *Aplysia* described earlier, it has been more difficult to link synaptic plasticity with specific examples of learning in vertebrates. However, one exciting and extensively studied candidate memory mechanism is the synaptic phenomenon termed long-term potentiation (LTP). This phenomenon is defined as a persistent increase in synaptic strength (as measured by the amplitude of the EPSP in a follower neuron) that can be induced rapidly by a brief burst of spike activity in the presynaptic afferents. The intense experimental interest in LTP is driven by the working hypothesis that this form of synaptic plasticity may participate in information storage in several brain regions. This section describes the properties of LTP and how it is studied, reviews its underlying mechanisms, and explores the possibility of linkages between LTP and learning and memory.

Long-Term Potentiation Occurs in a Variety of Neural Synapses

The first evidence that long-term modification of mammalian synapses could be induced by experimental means appeared in 1973, when Timothy Bliss and Terje Lomo demonstrated LTP in the hippocampus of the anesthetized rabbit. Brief, high-frequency stimulation (HFS) of the perforant-pathway input to the dentate gyrus produced a long-lasting enhancement of the extracellularly recorded field potential. Subsequent studies of nonanesthetized animals have shown that LTP can last for weeks or months. Originally thought to be unique to the mammalian hippocampal formation, LTP is now known to occur in the cerebellum, neocortical regions, subcortical regions such as the amygdala, mammalian peripheral nervous system, the arthropod neuromuscular junction, and the *Aplysia* sensorimotor synapse. It is important to bear in mind that no universal mechanism exists for inducing LTP. Indeed, different mechanisms may be used at the same synapse, depending on the experimental conditions. This chapter focuses primarily on the LTP at the synapse made by a pyramidal neuron in the CA3 region of the hippocampus to a pyramidal neuron in the CA1 region of the hippocampus.

Long-Term Potentiation at the CA3–CA1 Synapse

Within the hippocampus proper, by far the best studied synapse is that from the Schaffer collateral/commissural (Sch/com) fibers of the CA3 pyramidal cells to the CA1 pyramidal cells (Fig. 47.6). In fact, this is probably the most commonly studied synapse in the mammalian brain, due in part to its relatively simple circuitry and its laminar organization. These features make it possible to extract

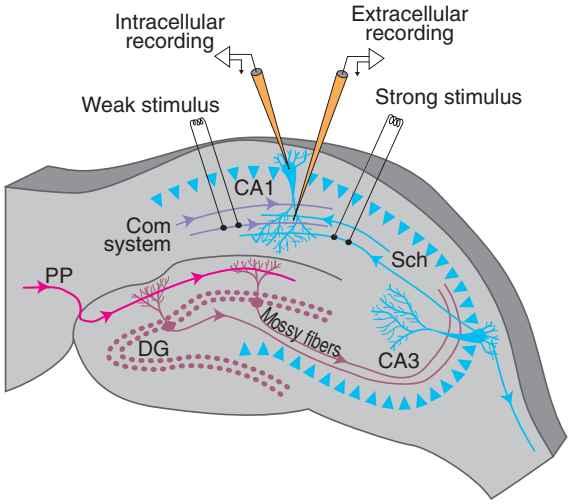


FIGURE 47.6 Schematic of a transverse hippocampal brain slice preparation from the rat. Two extracellular stimulating electrodes are used to activate two nonoverlapping inputs to pyramidal neurons of the CA1 region of the hippocampus. By suitably adjusting the current intensity delivered to the stimulating electrodes, different numbers of Schaffer collateral/commissural (Sch/com) axons can be activated. In this way, one stimulating electrode was made to produce a weak postsynaptic response and the other to produce a strong postsynaptic response. Also illustrated is an extracellular recording electrode placed in the stratum radiatum (the projection zone of the Sch/com inputs) and an intracellular recording electrode in the stratum pyramidale (the cell body layer). Also indicated is the mossy fiber projection from granule cells of the dentate gyrus (DG) to the pyramidal neurons of the CA3 region. *Adapted from Barrionuevo and Brown (1983).*

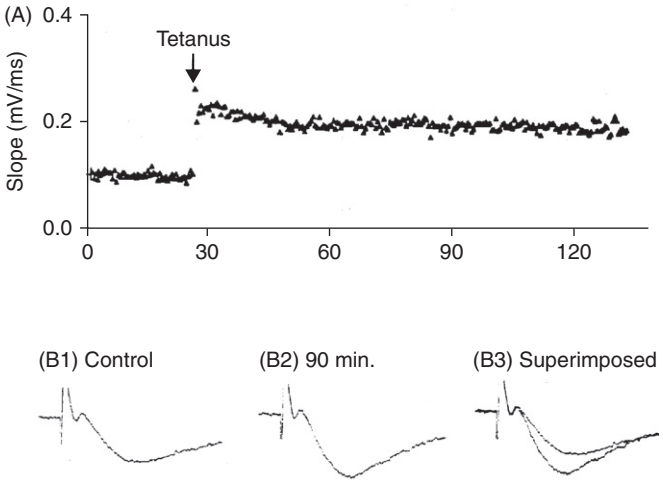


FIGURE 47.7 LTP at the CA3–CA1 synapse in the hippocampus. (A) Test stimuli are delivered repeatedly once every 10 s while the strength of the synaptic connection is monitored. Strength can be assessed by the *amplitude* of the extracellularly recorded EPSP or, as was done in this example, as the *slope* of the rising phase of the EPSP, which provides an accurate reflection of its strength. To induce LTP, two 1 s, 100 Hz tetani were delivered with a 20-s interval. Subsequent test stimuli produce enhanced EPSPs. The enhancement is stable and persists for at least 2 h. Examples of extracellulary recorded field EPSPs before (B1) and 90 min after the induction of LTP (B2). In B3 the traces from B1 and B2 are superimposed. *Modified from Nicoll et al. (1988).*

useful data from extracellular recordings, which are easier to perform than intracellular recordings and preferable for some purposes. Examples of LTP induction are illustrated in Figure 47.7B. The lower waveforms are extracellularly recorded field EPSPs recorded in the CA1 region in response to a single weak stimulation of the Sch/com pathway. Brief electric stimuli delivered to this pathway lead to the initiation of action potentials in the individual axons in the pathway. These action potentials then propagate to the synaptic terminals. The release of transmitter from the multiple afferent terminals produces a summated EPSP in the postsynaptic cell, which can be detected with an extracellular electrode. Test stimuli are delivered repeatedly at a low rate that produces stable EPSPs in the postsynaptic cell (Figs. 47.7A and B1). After a baseline period, a brief high-frequency tetanus is delivered. Subsequent test stimuli produce enhanced EPSPs (Fig. 47.7B2). The enhancement persists for many hours. Although the synaptic enhancement is stable after the tetanus, LTP, like heterosynaptic facilitation of the sensorimotor synapse in *Aplysia*, has multiple temporal domains. One domain is associated with an enhancement of the EPSP that persists for about 90 min. This form of LTP is referred to as early LTP (E-LTP). A second domain referred to as late LTP (L-LTP) is associated with synaptic enhancement that persists for periods of time greater than about 90 min. As described later, different mechanisms underlie the induction and maintenance of E- and L-LTP.

Properties of Long-Term Potentiation at the CA3–CA1 Synapse Include Cooperativity, Associativity, and Input Specificity

The CA3–CA1 synapses exhibit a form of LTP characterized by "classical" properties that have been termed "cooperativity," "associativity," and "input specificity" (Fig. 47.8) (Bliss & Collingridge, 1993; Brown, Ganong, Kairiss, & Keenan, 1990). These "classical properties" are actually different manifestations of the same underlying mechanism that is responsible for this type of LTP. Other less commonly studied types of LTP have different signatures.

Cooperativity refers to the fact that the probability of inducing LTP, or the magnitude of the resulting change, increases with the number of stimulated afferents. Weak HFS, which activates fewer afferents, often fails to induce LTP (Fig. 47.8A). In contrast, strong stimulation, which activates more afferents, produces LTP more reliably (Fig. 47.8B). Thus, the additional axons recruited by higher stimulation intensities "cooperate" to trigger LTP.

Associativity was shown in preparations in which two distinct axonal inputs converged onto the same postsynaptic target. Consider the interactions between two stimulus pathways, one termed the *weak* pathway with a

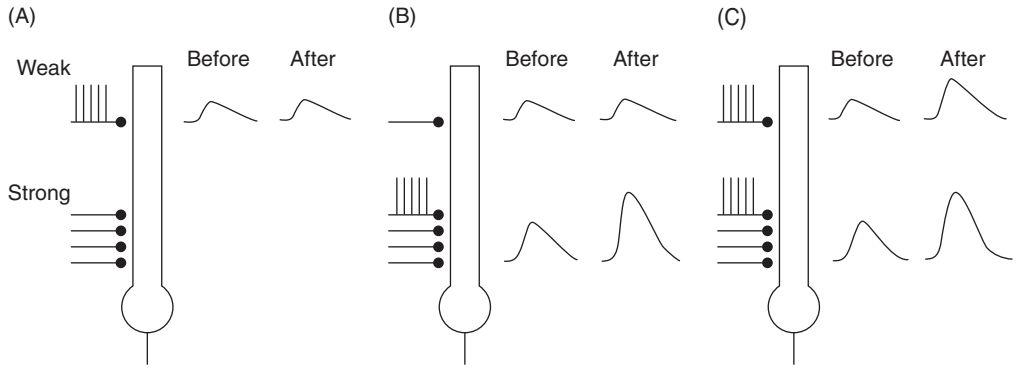


FIGURE 47.8 Features of LTP at CA3–CA1 synapses in the hippocampus. A single hippocampal pyramidal cell is shown receiving a weak and strong synaptic input. (A) Tetanic stimulation of the weak input alone does not cause LTP in that pathway (compare the EPSP before and after the tetanus). (B) Tetanic stimulus of the strong input alone causes LTP in the strong pathway, but not in the weak pathway. (C) Tetanic stimulation of both the weak and the strong pathway together causes LTP in both the weak and the strong pathway. *Modified from Nicoll et al. (1988).*

small number of stimulated afferents, and the other termed the *strong* pathway with a large number of stimulated afferents (Fig. 47.8C). HFS of the *weak* input by itself failed to produce LTP in that pathway unless this stimulation was paired with tetanic stimulation of the *strong* input. Thus, LTP was induced in a *weak* input only when its activity was associated with activity in the *strong* input.

Input specificity means that LTP is restricted to only the inputs that received the tetanic stimulation. An unstimulated *weak* pathway was not facilitated after the tetanus to the *strong* pathway (Fig. 47.8B).

A Hebbian Mechanism Explains the Properties of Long-Term Potentiation at the CA3–CA1 Synapse in the Hippocampus

How can these classical properties of LTP in the CA1 region of the hippocampus be explained? In the late 1940s, the Canadian psychologist Donald Hebb (1949) formulated a postulate regarding the conditions that cause synapses to change. His thinking proved to be influential and guided later experiments that probed the mechanisms behind LTP. According to Hebb's postulate:

> When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased. (p. 62)

In short, coincident activity in two synaptically coupled neurons was proposed to cause increases in the synaptic strength between them. Numerous modern interpretations of Hebb's postulate exist, but most are captured by the mnemonic: "Cells that fire together, wire together."

Could the classical properties of LTP all be consequences of synapses that obey a Hebbian rule? Possibly so if a critical amount of postsynaptic depolarization were a necessary condition for inducing LTP in active synapses. In this case, cooperativity would result when enough input fibers were stimulated to produce the critical amount of postsynaptic depolarization. Associativity would emerge from the fact that the strong input caused sufficient depolarization of the postsynaptic membrane during the presynaptic activity in the weak input. Input specificity would occur because LTP was induced only in those inputs to a neuron that were active at the same time that the cell was sufficiently depolarized by the strong input to that neuron. In other words, these classical phenomena could all be manifestations of a single underlying Hebbian mechanism at the CA3–CA1 synapse.

Not all forms of LTP are Hebbian, however. Examples of non-Hebbian LTP can be found at the mossy fiber–CA3 synapse in the hippocampus and at the parallel fiber–Purkinje cell synapse in the cerebellum. These results indicate that the classical properties of cooperativity, associativity, and input specificity are not universal.

Mechanisms for Induction, Expression, and Maintenance of Long-Term Potentiation

LTP Induction

It is currently thought that there are multiple mechanisms or at least multiple second-messenger pathways that can lead to persistent synaptic enhancement. Multiple mechanisms may also contribute to expression and maintenance.

CALCIUM IONS AND LTP

It is generally agreed that the induction of LTP depends on an increase in the intracellular concentration of calcium ions ($[Ca^{2+}]_i$) in some key compartments of pre- and/or postsynaptic cells (Bliss & Collingridge, 1993; Johnston, Williams, Jaffe, & Gray, 1992; Nicoll & Malenka, 1995). The exact role of calcium in the induction process depends on the particular form of LTP and the synaptic system. In the CA1 region of the hippocampus, LTP induction in the Sch/com synapse depends on changes in postsynaptic

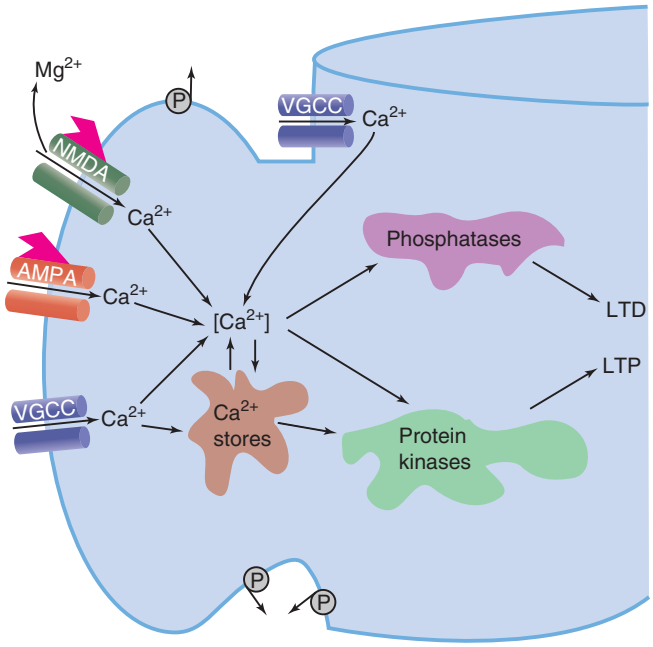


FIGURE 47.9 Events leading to some forms of LTP and LTD. The schematic depicts a postsynaptic spine with various sources of Ca^{2+}. The NMDA receptor channel complex admits Ca^{2+} only after depolarization removes the Mg^{2+} block in the presence of bound glutamate. Calcium may also enter through the ligand-gated AMPA receptor channel or voltage-gated calcium channels (VGCC), which may be located on the spine head or dendritic shaft. Calcium pumps (P), located on the spine head, neck, and dendritic shaft, are hypothesized to help isolate Ca^{2+} concentration changes in the spine head from those in the dendritic shaft.

$[Ca^{2+}]_i$ (Fig. 47.9). Two major pathways that have been studied extensively are implicated in some aspect of LTP induction: calcium influx through ionotropic GluRs, especially the *N*-methyl-D-aspartate receptor (NMDAR); and calcium influx through voltage-gated calcium channels (VGCCs).

NMDAR-DEPENDENT LTP

Recall that the classical form of LTP in the CA1 region of the hippocampus has properties that can be explained in terms of a Hebbian mechanism. For this form of LTP, considerable evidence shows a role for the NMDAR (Bliss & Collingridge, 1993). Numerous pharmacological studies have shown that competitive antagonists of NMDA, such as D-2-amino-5-phosphonopentenoic acid (D-AP5, also termed AP5 or APV) or NMDA ion channel blockers, such as the noncompetitive antagonist (+)-5-methyl-10, 11-dihydro-*5H*-dibenzo[*a,d*] cyclohepten-5,10-imine (MK-801), can prevent the induction of LTP.

The NMDAR has two properties that immediately suggest the nature of its role in LTP induction at Hebbian synapses (Bliss & Collingridge, 1993; Brown et al., 1990; Gustafsson & Wigstrom, 1988). First, NMDARs are permeable to Ca^{2+} (in addition to Na^+ and K^+). This property is significant because postsynaptic $[Ca^{2+}]_i$ plays a critical role in inducing NMDAR-dependent LTP. Second, the channel permeability is a function of both pre- and postsynaptic factors. Channel opening requires the neurotransmitter glutamate (or some related agonist) to bind to the NMDA site. This in turn requires presynaptic activity for glutamate release. At the usual resting membrane potential, the ionic channels of NMDARs are normally blocked by magnesium ions (Mg^{2+}), but this channel block is relieved by sufficient depolarization of the postsynaptic membrane. Thus, the NMDAR-mediated conductance is voltage dependent, allowing Ca^{2+} entry only when presynaptic release is combined with postsynaptic depolarization.

At this point, you should recall the distinction between the properties of the NMDAR and those of the AMPAR (α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid receptor) (see Chapters 8 and 10 for details), which are also found in the postsynaptic membrane. The AMPAR does not exhibit voltage-dependent Mg^{2+} block, has relatively lower Ca^{2+} permeability, and the AMPAR-mediated conductance is essentially voltage independent. Released glutamate can potentially act on both the AMPARs and the NMDARs associated with the membrane on the dendritic spine (Fig. 47.9). With this knowledge, one can easily envision a possible role for the AMPAR and NMDAR in a Hebbian modification. Nearly concurrent presynaptic activity (producing glutamate release and binding) and postsynaptic depolarization (relieving the Mg^{2+} block) allow Ca^{2+} influx into the dendritic spine of the postsynaptic neuron. The increased $[Ca^{2+}]_i$ in some critical region of the dendritic spine, presumably very close to the NMDAR, is thought to activate Ca^{2+}-dependent enzymes, such as CaM kinase II, which play a key role in LTP induction (Fig. 47.9).

In qualitative terms, one can understand how these molecular events could help account for the properties of cooperativity, associativity, and spatiotemporal specificity. Active synapses release glutamate, which can bind to the NMDAR, causing Ca^{2+} influx into dendritic spines on the postsynaptic cell. This Ca^{2+} influx acts locally and results in input-specific LTP. However, the Ca^{2+} influx occurs only when the synaptic input is strong enough to depolarize the postsynaptic membrane sufficiently to relieve the Mg^{2+} block, giving rise to cooperativity (Fig. 47.8). The depolarization itself is mediated in large part by the (voltage-independent) AMPARs, which are also colocalized on the dendritic spine (Fig. 47.9). Note that activity in a weak input by itself would not depolarize the postsynaptic cell sufficiently to relieve the Mg^{2+} block unless this activity were properly timed in relationship to activity in a strong input to the same cell. The combined depolarization of the two inputs gives rise to associativity and input specificity (Fig. 47.8).

NMDAR-INDEPENDENT LTP

Most of the preceding accounts of the Hebbian, NMDAR-dependent form of LTP apply only to certain synapses under some conditions, and even then it may be only one aspect of the story (Johnston et al., 1992; Nicoll & Malenka, 1995). In many synapses, LTP induction does not appear to require the NMDAR (Johnston et al., 1992; Teyler et al., 1994). Even within the hippocampus, some synapses exhibit NMDAR-independent forms of LTP. For example, in the presence of the competitive antagonist APV, even the Sch/com inputs to CA1 pyramidal neurons, which are known to exhibit the classical Hebbian form of LTP that relies on the NMDAR, can exhibit an NMDAR-independent type of LTP. The onset of NMDAR-independent LTP is relatively slow (20–30 min), exhibits input specificity, and is prevented by a blocker of L-type voltage-gated calcium channels (VGCCs). Thus, a distinction exists between "NMDA LTP" and "VGCC LTP." Other work has suggested that VGCCs are likely to be responsible for certain types of LTP in the CA3 region of the hippocampus and in the visual cortex.

The general case may be that both NMDAR-dependent and NMDAR-independent forms of LTP may coexist in the same brain region, among different classes of synaptic inputs onto the same postsynaptic neuron, and even among the same class of synaptic inputs to the same postsynaptic neuron.

INDUCTION OF LATE LTP

A common requirement for the induction of both early LTP and late LTP at the CA3–CA1 synapse is the elevation of levels of intracellular calcium in the postsynaptic (i.e., CA1) neuron. Additional steps are involved in the induction of L-LTP, however. As was the case for the induction of long-term facilitation of the sensorimotor synapse in *Aplysia*, activation of PKA appears to be necessary for the induction of L-LTP (Huang, Nguyen, Abel, & Kandel, 1996). However, in the CA1 neuron the cAMP pathway is engaged directly by the activation of a calcium-sensitive adenylyl cyclase rather than being activated by 5-HT as in *Aplysia*. (Transmitters that regulate cAMP in CA1 neurons can profoundly modulate LTP, however.) Thus, at the CA1 neuron elevated levels of calcium lead to activation of adenylyl cyclase, increased synthesis of cAMP, and activation of PKA. Also, like long-term facilitation in *Aplysia*, MAPK is necessary for the induction of L-LTP (Dineley et al., 2001). The mechanisms of activation of MAPK in CA1 neurons have not been elucidated in detail, but may involve PKA, PKC, or both acting together. The final steps in the induction of L-LTP involve a PKA- and MAPK-dependent phosphorylation of CREB and induction of CREB responsive genes. These genes encode the proteins underlying changes at the synapse (but see Box 47.1).

The mechanisms underlying the induction of long-term neuronal plasticity and memory seem to be highly conserved across species. NMDAR-dependent LTP has been characterized at the sensorimotor synapse of *Aplysia* (Lin & Glanzman, 1994). Formation of some long-term memories in *Drosophila* depends on both an increase in the activity of a transcriptional activator of the bZip family, dCREB2a, and a decrease in the activity of a related repressor, dCREB2b. Establishment of LTM in *Drosophila* therefore appears analogous to induction of *Aplysia* LTF, involving transcriptional activation by one bZip family member (CREB1 in *Aplysia*) and relief of

BOX 47.1

IS MEMORY MORE THAN CHANGES IN SYNAPTIC STRENGTH?

The search for the biological basis of learning and memory has led many of the leading neuroscientists of the twentieth century to direct their efforts to investigating the synapse. The focus of most recent work on LTP (and indeed the bulk of this chapter) has been to elucidate the mechanisms underlying changes in synaptic strength. Although changes in synaptic strength are certainly ubiquitous, they are not the exclusive means for the expression of neuronal plasticity associated with learning and memory. Both short-term and long-term sensitization and classical conditioning of defensive reflexes in *Aplysia* are associated with an enhancement of excitability of the sensory neurons in addition to changes in synaptic strength. Classical and operant conditioning of feeding behavior in *Aplysia* also leads to changes in excitability of neuron B51. Interestingly, in this case there is a bidirectional control. The excitability of B51 is increased by operant conditioning, whereas it is decreased by classical conditioning. Changes in excitability of sensory neurons in the mollusc *Hermissenda* are produced by classical conditioning. In vertebrates, classical conditioning of eye-blink reflexes produces changes in the excitability of cortical neurons. Eye-blink conditioning also produces changes in the spike afterpotential of hippocampal pyramidal neurons. Finally, as described in their original report on LTP, Bliss and Lomo found that the expression of LTP was also associated with an apparent enhanced excitability. For recent review see Mozzachoidi and Byrne (2010).

John H. Byrne

repression by another member (CREB2 in *Aplysia*). CREB activation also correlates with mammalian long-term memories. Interference with CREB function can inhibit long-term memories whereas overexpression of CREB can facilitate them.

LTP Expression

Up to this point, evidence related to early events in the causal chain that triggers the induction process has been emphasized. Another question follows naturally: What biochemical and biophysical changes incorporate this modification once it has been triggered? Most of the ideas about enhanced synaptic transmission concern either increased transmitter release or increased receptivity to released transmitter. The former entails presynaptic changes the latter, postsynaptic. Although some of the induction mechanisms discussed previously implicated a postsynaptic increase in $[Ca^{2+}]_i$, this does not necessarily imply that expression must also be postsynaptic. Ample evidence is available for ongoing two-way communication across the synaptic cleft, so a postsynaptic trigger could, in principle, give rise to a pre- and/or postsynaptic modification (see also Fig. 47.5). Extensive and seemingly conflicting accounts can be found in the literature regarding the nature of the changes responsible for the observed increase in synaptic efficacy following LTP induction. Some of the possibilities are illustrated schematically in Figure 47.10. However, regarding LTP at the CA3–CA1 synapse, the expression of LTP is associated with an increase in the number of functional AMPA receptors in the postsynaptic neuron (*mechanism No. 5* in Fig. 47.10; for review, see Malinow & Malenka, 2002). LTP at this synapse also involves phosphorylation of the AMPA receptor, which leads to an increased conductance that contributes to enhanced synaptic efficacy (*mechanism No. 4* in Fig. 47.10). Recent results indicate that a family of small transmembrane AMPA receptor regulatory proteins (TARPs) controls both AMPA receptor trafficking and channel gating. Finally, LTP at hippocampal and cortical excitatory synapses correlates with increases in the size of dendritic spines on the postsynaptic neurons (reviewed in Kasai, Fukuda, Watanabe, Hayashi-Takagi, & Noguchi, 2010).

In contrast to the predominant evidence for postsynaptic expression of LTP at the CA3–CA1 synapse (see, however, Enoki, Hu, Hamilton, & Fine, 2009), LTP at the mossy fiber–CA3 synapse appears to involve an increase in transmitter release.

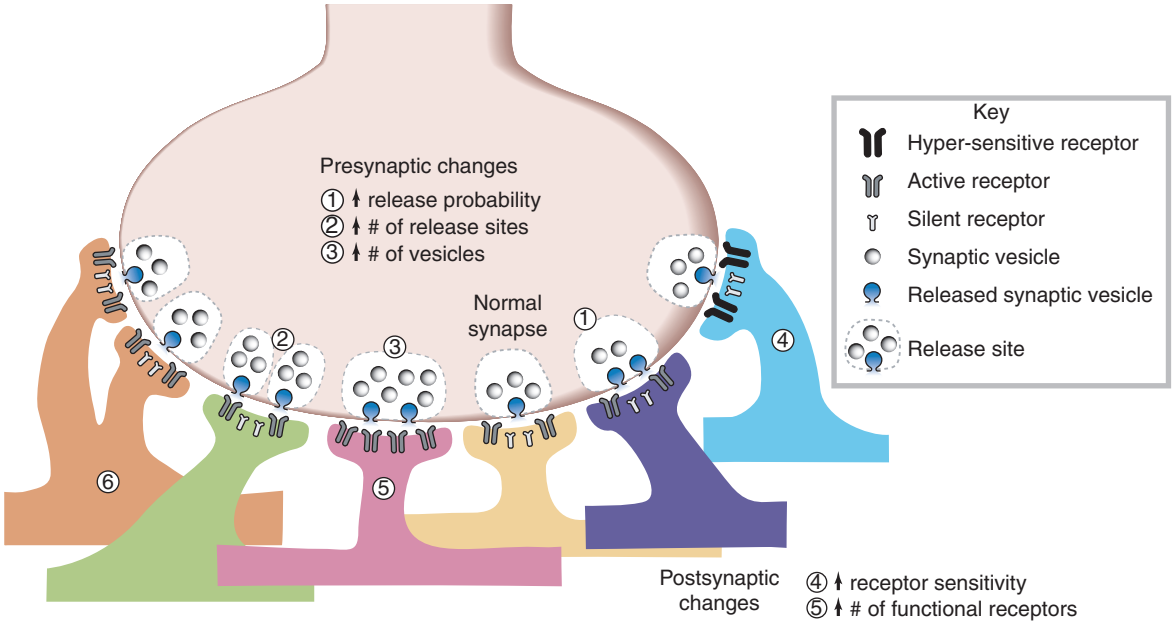


FIGURE 47.10 Schematic representation of possible loci for cellular changes involved in the enhancement of synaptic efficacy. The efficacy of a synapse can be potentiated through at least six mechanisms. First, there could be an increase in the fraction (release probability) of available presynaptic vesicles that undergo exocytosis. For example, in *mechanism 1*, two out of four available vesicles are released (i.e., 50% release probability), in contrast to the *normal synapse*, where only one out of four vesicles is released (i.e., 25% release probability). Second, there could be an increase in the number of release sites at the presynaptic neuron (*mechanism 2*). Third, the synapse could be potentiated through an increase in the number of vesicles available for release. For example, at a release site with eight vesicles, two of them will be exocytosed (instead of one) (*mechanism 3*), even if the release probability (25%) is the same as at the normal synapse. Fourth, there could be an increase in the sensitivity of the preexisting receptors to presynaptically released neurotransmitter or a greater conductance of the channel (*mechanism 4*). Fifth, there could be an increase in the number of functional receptors (illustrated as an increase from two active receptors at the normal synapse to four at the potentiated synapse; *mechanism 5*). Finally, the synapse could also be potentiated through coordinated presynaptic and postsynaptic morphological changes, such as the growth of new synaptic contacts between the same pair of neurons (*mechanism 6*). *Adapted from Wang, Ko, and Kelly (1997).*

LTP Maintenance

Regardless of the ultimate nature and locus of the modification that gives rise to LTP expression, the more general problem remains of how a synaptic change can endure over long periods of time in the face of constant molecular turnover.

MAINTENANCE OF EARLY LTP

The maintenance of E-LTP is due to the persistence of phosphorylation of substrate proteins involved in expression mechanisms (see earlier discussion). This phosphorylation, in turn, is regulated by the engagement of protein kinases and phosphatases. For E-LTP, autonomously active forms of PKC and CaM kinase II appear to be particularly important.

MAINTENANCE OF LATE LTP

The maintenance of persistent forms of LTP ultimately involves alterations in gene expression and changes in protein synthesis. High-frequency electrical stimulation in the rat hippocampus raises levels of specific mRNAs that encode transcription factors (e.g., Fos, zif268), cytoskeletal proteins (e.g., Arc), and signal transduction molecules such as CaM kinase II. Moreover, protein synthesis inhibitors block the late phase of LTP but not earlier phases. One interesting observation is that brain slices that received protein synthesis inhibitors just 2 h after HFS showed no decline in late LTP, indicating that there is a critical time window during which protein synthesis might be necessary to maintain long-term plasticity. One protein PKMζ appears both necessary and sufficient for maintenance of a least some forms of L-LTP and seems necessary for maintaining the increase in postsynaptic AMPA receptors (Sacktor, 2011).

Although the involvement of new protein synthesis is consistent with the data, it immediately raises the problem of synaptic input specificity. If neural activity ultimately affects gene expression in the nucleus, then the mRNAs and proteins produced in the somatic compartment could, in principle, travel to any synapse within the cell. The problem is how to modify only the appropriate synapses and maintain synapse specificity. One solution is for the specific kinase(s) activated locally during a tetanus to phosphorylate one or more substrates and that these substrates act to "tag" the synapse as one that has been potentiated. For example, a stimulus that activates one CA1 synaptic input sufficiently to induce L-LTP can lead to the induction of L-LTP by a weaker stimulus to a second synaptic input even when that second input is by itself sufficient to only induce E-LTP. This effect is limited to a time window where the second weaker stimulus occurs within an hour or so of the stronger one. These results indicate that the weaker stimulus can lead to the transient (~1 h) activation of a synaptic "tag" that can specificially "capture" "plasticity factors" induced by a strong stimulus to a different synapse (Redondo & Morris, 2011). A major goal of current LTP research is to identify the nature of the tag and the plasticity factors.

Support of a more permanent synaptic modification might require a self-perpetuating tag and/or a biochemical positive feedback loop that maintains synaptic strength (Hayer & Bhalla, 2005). One possibility is that CPEB, the prion-like molecule that has been implicated in *Aplysia* LTF (Miniaci et al., 2008), might help to maintain synthesis of PKMζ (Sacktor, 2011). Positive feedback based on network activity may also be important. Periodic reactivation of strengthened synapses may be necessary to reactivate mechanisms of L-LTP, maintaining synaptic strength for months or longer (Smolen, 2007; Wang & Tsien, 2006).

Links between Long-Term Potentiation and Learning

Long-term potentiation has properties that have long been considered necessary for the encoding and retrieval of information. Hebbian forms of LTP exhibit associativity, which appears to be a desirable property, and all forms of LTP appear to be well suited to rapid learning. One of the more important challenges entails linking LTP (and LTD, see later) to learning and memory. This task has proven rather difficult to achieve, although evidence is accumulating for a causal link between LTP and memory. For example, learning in rats was verified to induce hippocampal LTP (Whitlock, Heynen, Shuler, & Bear, 2006). Also, inhibition of PKMζ in the hippocampus *in vivo* was demonstrated to disrupt LTP maintenance and eliminate spatial learning. Instrumental and classically conditioned long-term memories are similarly erased by locally inhibiting PKMζ in different brain regions of rats and mice, from days to even weeks and months after training (Saktor, 2011; Serrano et al., 2008). These results have greatly strengthened the link between LTP and learning. However, inhibiting PKMζ does not erase all forms of memory (Serrano et al., 2008).

Additional promising work in this area has come from studies of genetically engineered knockout mice. The first generation of knockouts prevented the expression of some factor that was thought to be necessary for LTP, such as CaM kinase II. Although these studies were intriguing, they suffered from the fact that the consequences of a gene knockout might alter brain development and might not be confined to a specific part of the brain under study. Some of these problems have been overcome in a second generation of knockouts that have temporal as well as spatial specificity. For example, temporal and spatial control of gene expression can now be achieved using binary transgene systems such as tetracycline transactivating systems and Cre/LoxP recombination systems. The NMDAR gene in only CA1 pyramidal cells of the hippocampus has been

knocked out using the Cre/LoxP system (Tsien, Huerta, & Tonegawa, 1996). The results provide strong evidence in favor of the notion that NMDA receptor-dependent synaptic plasticity at CA3–CA1 synapses is required for the acquisition of spatial memory. Further spatial specificity can be obtained by using promoters that are specific for subfields of the hippocampus. For example, one of the kainate receptor subunits, KA-1, exhibits high levels of expression in the CA3 pyramidal cell layer. Using the KA-1 promoter and the α-CaM kinase II promoter, Nakashiba, Young, McHugh, Buhl, and Tonegawa (2008) drove expression of tetanus toxin specifically in CA3 neurons. Expression of the toxin inhibited transmitter release from CA3 neurons which was associated with a block of one-trial contextual learning. Similar approaches using cell-specific promoters to drive expression of toxins or their receptors have been used to examine the role of cerebellar circuits that mediate classical conditioning of eye-blink reflexes (for review, see Nakanishi, 2009). The molecular approaches described above hold tremendous promise for testing hypotheses of the functional role of various forms of LTP in specific brain regions as well as the role of specific neural circuits in mediating memory processes.

LONG-TERM DEPRESSION

LTD is believed by many to be the mechanism by which learning is encoded in the cerebellum (Ito, 2001), as well as a process whereby LTP could be reversed in the hippocampus and neocortex (Bear & Malenka, 1994). In the hippocampus, brief HFS (e.g., four trains of 10 shocks at 100 Hz) can induce classical LTP, whereas low-frequency stimulation (LFS) over longer periods (1 Hz for 10 min) can induce LTD.

Some forms of LTD appear to be mediated by the NMDAR and these forms appear to be due to dephosphorylation of AMPAR subunits, which decreases their conductance and hence the synaptic efficacy, as well as endocytosis of AMPARs, which decreases their surface number and hence the synaptic efficacy. In addition, NMDAR-independent forms of LTD exist. These often depend on activation of metabotropic glutamate receptors. For example, NMDAR-independent LTD in the parallel fiber input to Purkinje cells in the cerebellum is induced by the combined increase in intracellular levels of Ca^{2+} (by activity) in postsynaptic Purkinje cells and activation of metabotropic glutamate receptors produced by the release of glutamate from parallel fibers.

The form of LTD may vary in different brain regions and sometimes among different inputs to the same brain region. Even within the CA1 region, the Sch/com input may exhibit both NMDAR-dependent and NMDAR-independent forms of LTD.

Calcium levels in the dendritic spines appear to be a common locus for the induction of NMDAR-dependent LTP and LTD. For example, at the Sch/com input to CA1, both LTP and LTD can be blocked by injecting Ca^{2+} chelators into the postsynaptic cell. If both LTP and LTD are triggered by Ca^{2+} entry, then how are their induction processes different? Presumably, more Ca^{2+} influx occurs during an LTP-inducing HFS than during an LTD-inducing LFS. One formal molecular model developed by John Lisman incorporates this Ca^{2+}-dependent, bidirectional control of synaptic strength. In this model, high $[Ca^{2+}]_i$ activates a protein kinase that phosphorylates a protein causing LTP induction, whereas intermediate $[Ca^{2+}]_i$ activates a protein phosphatase that dephosphorylates this protein and causes LTD. The synaptic strength thus depends on which of these competing processes is most active, which in turn will be a function of the pattern of activity experienced by the cell (Fig. 47.9).

LTP and LTD can be readily reversed. If LFS is applied after a synapse is potentiated by HFS, the synaptic strength decreases. This phenomenon is called depotentiation. In contrast, if HFS is applied after a synapse is depressed by LFS, the synaptic strength increases. This phenomenon is called dedepression. Accumulating evidence indicates that the mechanisms of depotentiation differ from that of LTD, and the mechanisms of dedepression differ from that of LTP.

Summary

The understanding of LTP and LTD is evolving rapidly (Bliss, Collingridge, & Morris, 2007). For example, whereas the *N*-methyl-D-aspartate receptor was once the pivotal focus of long-term potentiation research, it is now clear that other mechanisms should be considered. It has also become evident that HFS is but one end of a spectrum of stimulations that can induce synaptic changes. Lower stimulation frequencies can induce long-term depression, which may share some common molecular mechanisms with LTP. Finally, several stages in the maintenance of LTP have been identified, and probably more will be found. A remaining challenge is to clarify the varieties of LTP and LTD mechanisms and to demonstrate their functional significance by establishing convincing links to the encoding and retrieval of information.

HOW DOES A CHANGE IN SYNAPTIC STRENGTH STORE A COMPLEX MEMORY?

The relationship between the synaptic changes and the behavior of conditioned reflexes can be straightforward because the locus for the plastic change is part of the mediating circuit. Thus, the change in the strength of the

sensorimotor synapse in *Aplysia* can be related to the memory for sensitization (e.g., Fig. 47.2A1). However, the idea that an increase in synaptic strength leads to an enhanced behavioral response, and a decrease in synaptic strength leads to a decreased behavioral response, can be misleading. For example, a decrease in synaptic strength in a postsynaptic neuron that exerts an inhibitory action can be translated into an enhanced behavioral response. Indeed, in the parallel fiber-to-Purkinje cell connection in the cerebellum, such a disinhibition is precisely the mechanism that has been proposed to contribute to classical conditioning of the eye-blink reflex (see Chapter 48 for a discussion of this form of conditioning). Nevertheless, for relatively simple reflex systems in which the circuit is well understood, it is possible to directly relate a change in synaptic strength to learning. However, in most other examples of memory, it is considerably less clear how the synaptic changes are induced and, once induced, how the information is retrieved. This is especially true in memory systems that involve the storage of information for patterns, facts, and events. Neurobiologists have turned to artificial neural circuits to gain insights into these issues.

A simple network that can store and "recognize" patterns is illustrated in Figure 47.11. The network is artificial, but nevertheless is inspired by actual circuitry in the CA3 region of the hippocampus. In this example, six different input projections make synaptic connections with the dendrites of each of six postsynaptic neurons (Fig. 47.11A). The postsynaptic neurons serve as the output of the network. Input projections can carry multiple types of patterned information, and these patterns can be complex. In order to simplify the present discussion, consider that the particular input pathway in Figure 47.11A carries information regarding the pattern of neural activity induced by a single brief flash of a spatial pattern of light. For example, activity in the top pathway (line *a*) might represent light falling on the temporal region of the retina, whereas activity in the pathway on the bottom (line *f*) might represent light falling on the nasal region of the retina. Thus, the spatial pattern of an image falling upon the retina could be reconstructed from the pattern of neuronal activity over the *n* (in this case 6) input projections to the network.

Three aspects of the circuit endow it with the ability to store and retrieve patterns. First, each of the input lines makes a sufficiently strong connection with its corresponding postsynaptic cell to activate it reliably. Second, each output cell (*z* to *u*) sends an axon collateral that makes an excitatory connection with itself as well as the other five output cells. This pattern of synaptic connectivity leads to a network of 36 synapses (42 including the 6 input synapses). Third, each of the 36 synaptic connections is modifiable through an LTP-like mechanism (see earlier discussion). Specifically, the strength of a particular synaptic connection is initially weak, but it will increase if the presynaptic *and* postsynaptic neurons are active at the same time. The circuit configuration with the embedded synaptic "learning rule" leads to an autoassociation or autocorrelation matrix. The autoassociation is derived from the fact that the output is fed back to the input where it associates with itself.

Now consider the consequences of presenting the patterned input to the network of Figure 47.11A. The input pattern will activate the six postsynaptic cells in such a way as to produce an output pattern that will be a replica of the input pattern. In addition, the pattern will induce changes in the synaptic strength of the active synapses in the network. For example, synapse 3 will be strengthened because the postsynaptic cell, cell *z*, and the presynaptic cell, cell *x*, will be active at the same time. Note also that synapses 1, 5, and 6 will be strengthened as well. This occurs because these input pathways to cell *z* are also active. Thus, all synapses that are active at the same time as cell *z* will be strengthened. When the pattern is presented again as in Figure 47.11B, the output of the cell will not only be governed by the input, but also by the feedback connections, a subset of which were strengthened (Fig. 47.11B, filled synapses) by the initial presentation of the stimulus. Thus, for output cell *z*, a component of its activity will be derived from input *a*, but components will also come from synapses 1, 3, 5, and 6. If each of the initially strong and newly modified synapses is assumed to contribute equally to the firing of output cell *z*, the activity would be five times greater than the activity produced by input *a* before the learning. After learning, the output is an amplified version of the input but the basic features of the pattern are preserved.

Note that the "memory" for the pattern does not reside in any one synapse or in any one cell. Rather, it is *distributed* throughout the network at multiple sites. The properties of these types of autoassociation networks have been examined by James Anderson, Teuvo Kohonen, David Marr, Edmond Rolls, David Wilshaw, and their colleagues and found to exhibit a number of phenomena that would be desirable for a biological recognition memory system. For example, such networks exhibit pattern completion. If a partial input pattern is presented, the autoassociation network can complete the pattern in the sense that it can produce an output that is approximately what is expected for the full input pattern. Thus, any part of the stored pattern can be used as a cue to retrieve the complete pattern. For the example of Figure 47.11, the input pattern was {101011}. This pattern led to an output pattern of {505055}. If the input pattern is degraded to {101000}, then neurons *z* and *x* each receive two less recurrent inputs, and neurons *v* and *u* each receive two less recurrent inputs and lose their direct input. Nevertheless neurons *z*, *x*, *v*, and *u* each still receive two recurrent inputs and neurons *z* and *x* receive direct inputs. Therefore, these four neurons fire,

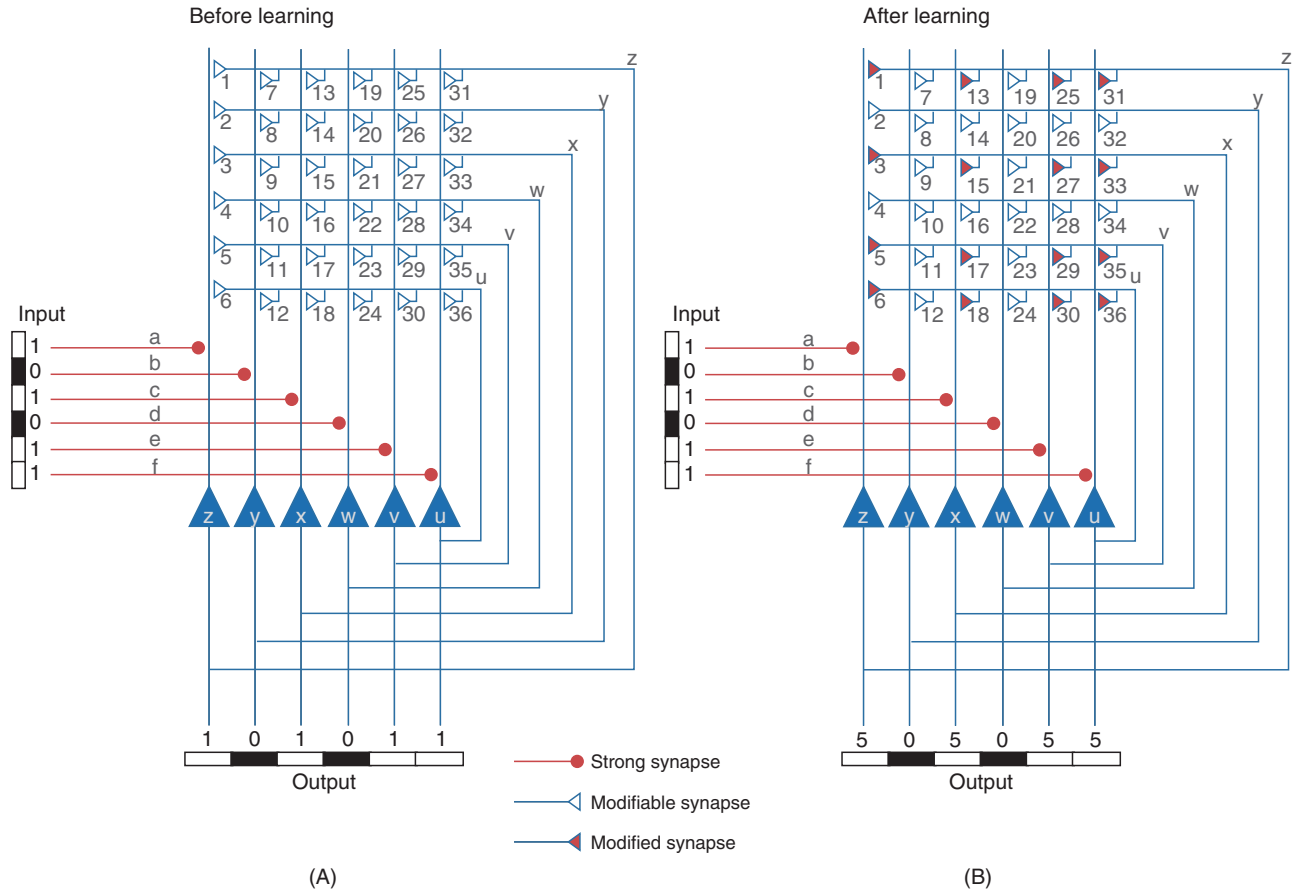


FIGURE 47.11 Autoassociation network for recognition memory. The artificial circuit consists of six input pathways that make strong connections to each of six output neurons. The output neurons have axon collaterals that make synaptic connections (numbered 1–36) with each of the output cells. (A) A pattern represented by activity in the input lines or axons (a, b, c, d, e, f) is presented to the network. A 1 represents an active axon (e.g., a spike), whereas a 0 represents an inactive axon. The input pathways make strong synapses (•) with the postsynaptic output cells. Thus, output cells (u, v, w, x, y, z) generate a pattern that is a replica of the input pattern. The collateral synapses were initially weak and do not contribute to the output. Nevertheless, the activity in the collaterals that occurred in conjunction (assume minimal delays within the circuit) with the input pattern led to a strengthening of a subset of the 36 synapses. (B) A second presentation of the input produces an output pattern that is an amplified, but an otherwise intact, replica of the input. An incomplete input pattern can be used as a cue to retrieve the complete pattern.

and the output pattern is {303022}. Compared to the original output pattern {505055}, some decrease in the strength of firing occurs, but the basic pattern is preserved. This preservation is a simple example of the phenomenon known as graceful degradation, which is exhibited by autoassociative networks. With graceful degradation, the network can still function if some of the input connections or postsynaptic cells are lost. This property arises from the distributed representation of the memory within the circuit.

Summary

The concept of distributed representation of memory crosses multiple levels of organization of memory systems. Multiple brain systems are involved in memory, and memory is distributed among synapses in a particular memory circuit (Fig. 47.11). Also, memory at any one synapse is represented by multiple cellular changes (Figs. 47.2, 47.3, and 47.9). The reductionist approaches described in this chapter have provided key insights into cellular memory mechanisms. In the near future, a major experimental question to be answered is the extent to which the mechanisms for learning are common both within any one animal and between different species. Although many common features are emerging, there seem to be some differences. Thus, it will be important to understand the extent to which specific mechanisms are used selectively for one type of learning and not another. Irrespective of the particular example of learning and memory that is analyzed, whether it be simple or complex, it will be important to pay attention to three major details: details of the circuit interactions, details of the learning rule, and details of the intrinsic biophysical properties of the neurons within the circuit.

References

Antonov, I., Kandel, E. R., & Hawkins, R. D. (2010). Presynaptic and postsynaptic mechanisms of synaptic plasticity and metaplasticity during intermediate-term memory formation in *Aplysia*. *Journal of Neuroscience, 30,* 5781–5791.

Antonov, I., Antonova, I., Kandel, E. R., & Hawkins, R. D. (2003). Activity-dependent presynaptic facilitation and Hebbian LTP are both required and interact during classical conditioning in *Aplysia*. *Neuron, 37,* 135–147.

Antzoulatos, E. G., & Byrne, J. H. (2007). Long-term sensitization training produces spike narrowing in *Aplysia* sensory neurons. *Journal of Neuroscience, 27,* 676–683.

Bailey, C. H., & Kandel, E. R. (2008). Synaptic remodeling, synaptic growth and the storage of long-term memory in *Aplysia*. *Progress in Brain Research, 169,* 179–198.

Barrionuevo, G., & Brown, T. H. (1983). Associative long-term potentiation in hippocampal slices. *Proceedings of the National Academy of Sciences of the United States of America, 80,* 7347–7351.

Bear, M. F., & Malenka, R. C. (1994). Synaptic plasticity: LTP and LTD. *Current Opinion in Neurobiology, 4,* 389–399.

Bliss, T. V. P., & Collingridge, G. L. (1993). A synaptic model of memory: Long-term potentiation in the hippocampus. *Nature* (Lond.), *361,* 31–39.

Bliss, T. V. P., Collingridge, G. L., & Morris, R. (2007). Synaptic plasticity in the hippocampus. In P. Anderson, R. Morris, D. Amaral, T. Bliss, & J. O'Keefe (Eds.), *The hippocampus book* (pp. 343–474). New York: Oxford University Press.

Brown, T. H., Ganong, A. H., Kairiss, E. W., & Keenan, C. L. (1990). Hebbian synapses: Biophysical mechanisms and algorithms. *Annual Review of Neuroscience, 13,* 475–512.

Byrne, J. H. (1987). Cellular analysis of associative learning. *Physiological Reviews, 67,* 329–439.

Byrne, J. H., & Kandel, E. R. (1996). Presynaptic facilitation revisited: State- and time-dependence. *Journal of Neuroscience, 16,* 425–435.

Cai, D., Chen, S., & Glanzman, D. L. (2008). Postsynaptic regulation of long-term facilitation in *Aplysia*. *Current Biology, 18,* 920–925.

Cai, D., Pearce, K., Chen, S., & Glanzman, D. L. (2011). Protein kinase M maintains long-term sensitization and long-term facilitation in *Aplysia*. *Journal of Neuroscience, 31,* 6421–6431.

Carew, T. J. (2000). *Behavioral neurobiology*. Sunderland, MA: Sinauer.

Chain, D. G., Casadio, A., Schacher, S., Hegde, A. N., Valbrun, M., Yamamoto, N., … Schwartz, J. H. (1999). Mechanisms for generating the autonomous cAMP-dependent protein kinase required for long-term facilitation in *Aplysia*. *Neuron, 22,* 147–156.

Choi, Y. B., Li, H. L., Kassabov, S. R., Jin, I., Puthanveettil, S. V., Karl, K. A., … Kandel, E. R. (2011). Neurexin-neuroligin transsynaptic interaction mediates learning-related synaptic remodeling and long-term facilitation in *Aplysia*. *Neuron, 70,* 468–481.

Cleary, L. J., Lee, W. L., & Byrne, J. H. (1998). Cellular correlates of long-term sensitization in *Aplysia*. *Journal of Neuroscience, 18,* 5988–5998.

Dineley, K. T., Weeber, E. J., Atkins, C., Adams, J. P., Anderson, A. E., & Sweatt, J. D. (2001). Leitmotifs in the biochemistry of LTP induction: Amplification, integration and coordination. *Journal of Neurochemistry, 77,* 961–971.

Enoki, R., Hu, Y. L., Hamilton, D., & Fine, A. (2009). Expression of long-term plasticity at individual synapses in hippocampus is graded, bidirectional, and mainly presynaptic: Optical quantal analysis. *Neuron, 62,* 242–253.

Fioravante, D., & Byrne, J. H. (2011). Protein degradation and memory formation. *Brain Research Bulletin, 85,* 14–20.

Fioravante, D., Antzoulatos, E. G., & Byrne, J. H. (2008). Sensitization and habituation: Invertebrate. In J. D. Sweatt (Ed.), *Volume 4 of Learning and memory: A comprehensive reference* (J.H. Byrene Editor), (Vols. 4, pp. 31–51). Oxford: Elsevier Science Limited.

Fioravante, D., Liu, R. Y., Netek, A. K., Cleary, L. J., & Byrne, J. H. (2007). Synapsin regulates basal synaptic strength, synaptic depression and serotonin-induced facilitation of sensorimotor synapses in *Aplysia*. *Journal of Neurophysiology, 98,* 3568–3580.

Ghirardi, M., Montarolo, P. G., & Kandel, E. R. (1995). A novel intermediate stage in the transition between short- and long-term facilitation in the sensory to motor neuron synapse of *Aplysia*. *Neuron, 14,* 413–420.

Guan, Z., Giustetto, M., Lomvardas, S., Kim, J. H., Miniaci, M. C., Schwartz, J. H., … Kandel, E. R. (2002). Integration of long-term-memory-related synaptic plasticity involves bidirectional regulation of gene expression and chromatin structure. *Cell, 111,* 483–493.

Gustafsson, B., & Wigstrom, H. (1988). Physiological mechanisms of long-term potentiation. *Trends in Neuroscience, 11,* 156–162.

Hawkins, R. D., Abrams, T. W., Carew, T. J., & Kandel, E. R. (1983). A cellular mechanism of classical conditioning in *Aplysia*: Activity-dependent amplification of presynaptic facilitation. *Science, 219,* 400–405.

Hayer, A., & Bhalla, U. S. (2005). Molecular switches at the synapse emerge from receptor and kinase traffic. *PLoS Computational Biology, 1,* 137–154.

Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley (Interscience).

Hu, J. Y., Wu, F., & Schacher, S. (2006). Two signaling pathways regulate the expression and secretion of a neuropeptide required for long-term facilitation in *Aplysia*. *Journal of Neuroscience, 26,* 1026–1035.

Huang, Y. -Y., Nguyen, P. V., Abel, T., & Kandel, E. R. (1996). Long-lasting forms of synaptic potentiation in the mammalian hippocampus. *Learning & Memory, 3,* 74–85.

Ito, M. (2001). Cerebellar long-term depression: Characterization, signal transduction, and functional roles. *Physiological Reviews, 81,* 1143–1195.

Jin, I., Kandel, E. R., & Hawkins, R. D. (2011). Whereas short-term facilitation is presynaptic, intermediate-term facilitation involves both presynaptic and postsynaptic protein kinases and protein synthesis. *Learning & Memory, 18,* 96–102.

Johnston, D., Williams, D., Jaffe, D., & Gray, R. (1992). NMDA-receptor-independent long-term potentiation. *Annual Review of Physiology, 54,* 489–505.

Kandel, E. R. (2001). The molecular biology of memory storage: A dialogue between genes and synapses. *Science, 294,* 1030–1038.

Kasai, H., Fukuda, M., Watanabe, S., Hayashi-Takagi, A., & Noguchi, J. (2010). Structural dynamics of dendritic spines in memory and cognition. *Trends in Neurosciences, 33,* 121–129.

Kim, J. H., Udo, H., Li, H. L., Youn, T. Y., Chen, M., Kandel, E. R., & Bailey, C. H. (2003). Presynaptic activation of silent synapses and growth of new synapses contribute to intermediate and long-term facilitation in *Aplysia*. *Neuron, 40,* 151–165.

Leal, K., & Klein, M. (2009). Direct enhancement of presynaptic calcium influx in presynaptic facilitation at *Aplysia* sensorimotor synapses. *Molecular and Cellular Neurosciences, 41,* 247–257.

Lechner, H. A., & Byrne, J. H. (1998). New perspectives on classical conditioning: A synthesis of Hebbian and Non-Hebbian mechanisms. *Neuron, 20,* 355–385.

Levenson, J., Endo, S., Kategaya, L. S., Fernandez, R. I., Brabham, D. G., Chin, J., … Eskin, A. (2000). Long-term regulation of neuronal high-affinity glutamate and glutamate uptake in *Aplysia*. *Proceedings of the National Academy of Sciences, 97,* 12858–12863.

Lin, X. Y., & Glanzman, D. L. (1994). Hebbian induction of long-term potentiation of *Aplysia* sensorimotor synapses: Partial requirement for activation of an NMDA-related receptor. *Proceedings Biological Sciences, 22,* 215–221.

Liu, R. Y., Cleary, L. J., & Byrne, J. H. (2011). The requirement for enhanced CREB1 expression in consolidation of long-term synaptic facilitation and long-term excitability in sensory neurons of *Aplysia*. *Journal of Neuroscience, 31,* 6871–6879.

Malinow, R., & Malenka, R. C. (2002). AMPA receptor trafficking and synaptic plasticity. *Annual Review of Neuroscience, 25,* 103–126.

Miniaci, M. C., Kim, J. H., Puthanveettil, S. V., Si, K., Zhu, H., Kandel, E. R., & Bailey, C. H. (2008). Sustained CPEB-dependent local protein synthesis is required to stabilize synaptic growth for persistence of long-term facilitation in *Aplysia*. *Neuron, 59,* 1024–1036.

Miyamoto, E. (2006). Molecular mechanism of neuronal plasticity: induction and maintenance of long-term potentiation in the hippocampus. *Journal of Pharmacological Sciences, 100,* 433–442.

Mozzachiodi, R., & Byrne, J. H. (2010). More than synaptic plasticity: Role of nonsynaptic plasticity in learning and memory. *Trends in Neurosciences, 33,* 17–26.

Nakanishi, S. (2009). Genetic manipulation study of information processing in the cerebellum. *Neuroscience, 162,* 723–731.

Nakashiba, T., Young, J. Z., McHugh, T. J., Buhl, D. L., & Tonegawa, S. (2008). Transgenic inhibition of synaptic transmission reveals role of CA3 output in hippocampal learning. *Science, 319,* 1260–1264.

Nicoll, R. A., Kaur, J. A., & Malenka, R. C. (1988). The current excitement of long-term potentiation. *Neuron, 1,* 97–103.

Nicoll, R. A., & Malenka, R. C. (1995). Contrasting properties of two forms of long-term potentiation in the hippocampus. *Nature (Lond.), 377,* 115–118.

Rachlin, H. (1991). *Introduction to modern behaviourism* (3rd ed.). New York: Freesman.

Redondo, R. L., & Morris, R. G. M. (2011). Making memories last: the synaptic tagging and capture hypothesis. *Nature Reviews Neuroscience, 12,* 17–30.

Roberts, A. C., & Glanzman, D. L. (2003). Learning in *Aplysia*: looking at synaptic plasticity from both sides. *Trends in Neuroscience, 26,* 662–670.

Sacktor, T. C. (2011). How does PKMzeta maintain long-term memory? *Nature Reviews Neuroscience, 12,* 9–15.

Serrano, P., Friedman, E. L., Kenney, J., Taubenfeld, S. M., Zimmerman, J. M., Hanna, J, … Fenton, A. A. (2008). PKMzeta maintains spatial, instrumental, and classically conditioned long-term memories. *PLoS Biology, 6,* 2698–2706.

Sharma, S. K., & Carew, T. J. (2004). The roles of MAPK cascades in synaptic plasticity and memory in *Aplysia*: Facilitatory effects and inhibitory constraints. *Learning & Memory, 11,* 373–378.

Si, K., Choi, Y. B., White-Grindley, E., Majumdar, A., & Kandel, E. R. (2010). *Aplysia* CPEB can form prion-like multimers in sensory neurons that contribute to long-term facilitation. *Cell, 140,* 421–435.

Smolen, P. (2007). A model of late long-term potentiation simulates aspects of memory maintenance. *PLoS ONE, 2,* e445.

Sutton, M. A., Masters, S. E., Bagnall, M. W., & Carew, T. J. (2001). Molecular mechanisms underlying a unique intermediate phase of memory in *Aplysia*. *Neuron, 31,* 143–154.

Teyler, T. J., Cavus, I., Coussens, C., DiScenna, P., Grover, L., Lee, Y. P., & Little, Z. (1994). Multideterminant role of calcium in hippocampal synaptic plasticity. *Hippocampus, 4,* 623–634.

Tsien, J. Z., Huerta, P. T., & Tonegawa, S. (1996). The essential role of hippocampal CA1 NMDA receptor-dependent synaptic plasticity in spatial memory. *Cell, 87,* 1317–1326.

Villareal, G., Li, Q., Cai, D., & Glanzman, D. L. (2007). The role of rapid, local, postsynaptic protein synthesis in learning-related synaptic facilitation in *Aplysia*. *Current Biology, 17,* 2073–2080.

Wainwright, M. L., Byrne, J. H., & Cleary, L. J. (2004). Dissociation of morphological and physiological changes associated with long-term memory in *Aplysia*. *Journal of Neurophysiology, 92,* 2628–2632.

Walters, E. T., & Byrne, J. H. (1983). Associative conditioning of single sensory neurons suggests a cellular mechanism for learning. *Science, 219,* 405–408.

Wang, J. H., Ko, G. Y., & Kelly, P. T. (1997). Cellular and molecular basis of memory: Synaptic and neuronal plasticity. *Journal of Clinical Neurophysiology, 14,* 264–293.

Wang, H., Hu, Y., & Tsien, J. Z. (2006). Molecular and systems mechanisms of memory consolidation and storage. *Progress in Neurobiology, 79,* 123–135.

Whitlock, J. R., Heynen, A. J., Shuler, M. G., & Bear, M. F. (2006). Learning induces long-term potentiation in the hippocampus. *Science, 313,* 1093–1097.

Zhang, F., Endo, S., Cleary, L. J., Eskin, A., & Byrne, J. H. (1997). Role of transforming growth factor-b in long-term facilitation in *Aplysia*. *Science, 275,* 1318–1320.

Zhao, Y., & Klein, M. (2002). Modulation of the readily releasable pool of transmitter and of excitation-secretion coupling by activity and by serotonin at *Aplysia* sensorimotor synapses in culture. *Journal of Neuroscience, 22,* 10671–10679.

Suggested Readings

Anderson, P., Morris, R, Amaral, D., Bliss, T., & O'Keefe, J. (Eds.), (2007). *The hippocampus book*. New York: Oxford University Press.

Byrne, J. H. (1997). Learning and memory. In J. H. Byrne (Ed.), *Neuroscience online: An electronic textbook for the neurosciences*. http://nba.uth.tmc.edu/neuroscience/ Department of Neurobiology and Anatomy, The University of Texas Medical School at Houston (UTHealth) © 1997.

Byrne, J. H. (Ed.), (2009). *Concise learning and memory—the editor's selection*. Oxford: Elsevier.

Dudai, Y. (2002). *Memory from a to z: Keywords, concepts, and beyond*. Oxford: Oxford University Press.

Kandel, E. R. (2007). *In search of memory: The emergence of a new science of mind*. New York: W. W. Norton & Company.

Squire, L. R., & Kandel, E. R. (1999). *Memory: From mind to molecules*. New York: Henry Holt and Company.

Sweatt, J. D. (2010). *Mechanisms of memory*. Amsterdam: Academic Press.

John H. Byrne

This page intentionally left blank

CHAPTER

48

Learning and Memory: Brain Systems

INTRODUCTION

The topic of memory systems can be usefully framed by the question of whether a specific memory can be localized or whether it is distributed throughout the brain. Even early on there was evidence that the answer might lie somewhere between these extremes. In the 1920s, the psychologist Karl Lashley conducted a series of experiments in which he carefully damaged various areas in the cerebral cortex of rats that had learned a route through a simple maze. He found that the total amount of cortex damaged correlated with the severity of memory impairment, irrespective of the specific location of the brain damage. Lashley therefore concluded that memory was distributed throughout the brain. Yet a short time later, the neurosurgeon Wilder Penfield found that when he stimulated various areas of the brains of awake epileptic patients (in order to identify functional areas deemed too important to be removed during surgery to treat the epilepsy), stimulation in some brain regions led the patient to experience specific memories. He reasoned that, if focal stimulations bring specific memories to mind, then memories might be localized in specific brain regions. The psychologist Donald Hebb offered reconciliation by suggesting that thoughts and memories were supported by "cell assemblies," or networks of neurons, and that experience resulted in changes in the connections between cells (Hebb, 1949). Thus, to the extent that a change at an individual juncture between cells (i.e., a synapse) recorded experience, memory was localized. Yet to the extent that full memories resulted from a reactivation of entire cell assemblies, memory was distributed.

The previous chapter discussed how synapses can be strengthened and weakened, and the current chapter considers how this synaptic plasticity supports behavior through memory systems, distributed cell assemblies involving a large number of interconnected neurons. A memory system is most usefully defined in both psychological and anatomical terms. That is, a memory system is most clearly identified by both a distinct neural circuitry and a unique set of operating characteristics. Memory systems are not separated according to stimulus modality (e.g., auditory vs. visual modality) or response modalities (e.g., manual vs. verbal responses). Instead, the critical distinctions involve how the biology of each system's anatomy and physiology support a particular type of memory representation in computational, psychological, and behavioral terms.

The present chapter focuses on individual structures that play central roles in each of the different systems and includes separate sections on the hippocampus, striatum, cerebellum, amygdala, and cerebral cortex (Fig. 48.1). Each of these structures anchors a much larger network of distributed brain areas that work in concert to support particular forms of memory. For example, the hippocampal memory system involves not only the hippocampus but also areas in the adjacent parahippocampal region and is connected to many distributed unimodal and polymodal neocortical areas. These areas normally work together to support conscious recollection of facts and events, a capacity termed declarative memory. Even though declarative memory depends on very specialized information processing and plasticity within the hippocampus, a complicated and distributed circuitry outside the hippocampus is also required for these alterations to emerge as the experience of remembering. Importantly, the cerebral cortex plays a role in many types of memory, and the present chapter will discuss

© 2013 Elsevier Inc. All rights reserved.

both its role as a part of multiple memory systems and the specific memory functions it supports outside of those systems. Finally, memory systems work in concert to support behavior in the course of one's day to day activities, and this issue will be considered in the final section of the chapter.

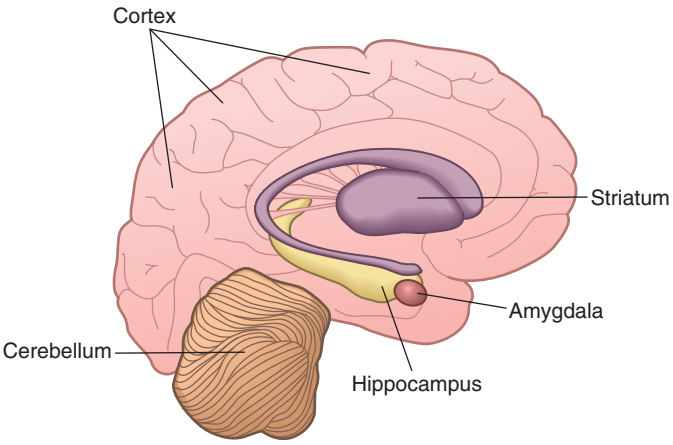


FIGURE 48.1 Drawing of the human brain showing components of the major memory systems.

Hippocampus

Anatomy

The hippocampal memory system includes the hippocampus (defined here as the CA fields, dentate gyrus, and subiculum) and the entorhinal, perirhinal, and postrhinal cortices in the adjacent parahippocampal region (the postrhinal cortex is referred to as the parahippocampal cortex in primates; Fig. 48.2; Burwell, Witter, & Amaral, 1995; Suzuki, 1996). The anatomy and circuitry of these regions, especially the hippocampus, are largely conserved across mammalian species (Manns & Eichenbaum, 2007). The parahippocampal region serves as a convergence site for input from cortical association areas and mediates the distribution of cortical afferents to the hippocampus. These parahippocampal cortical areas are interconnected and send major efferents to multiple subdivisions of the hippocampus. Within the hippocampus, an intricate pattern of connectivity mediates a large network of associations (Fig. 48.2; Amaral & Witter, 2004), and these connections support forms of long term potentiation that could participate in the rapid coding of novel conjunctions of information (see Chapter 47). The outcomes of hippocampal processing are directed back to the adjacent cortical areas in the parahippocampal region, and the outputs of that

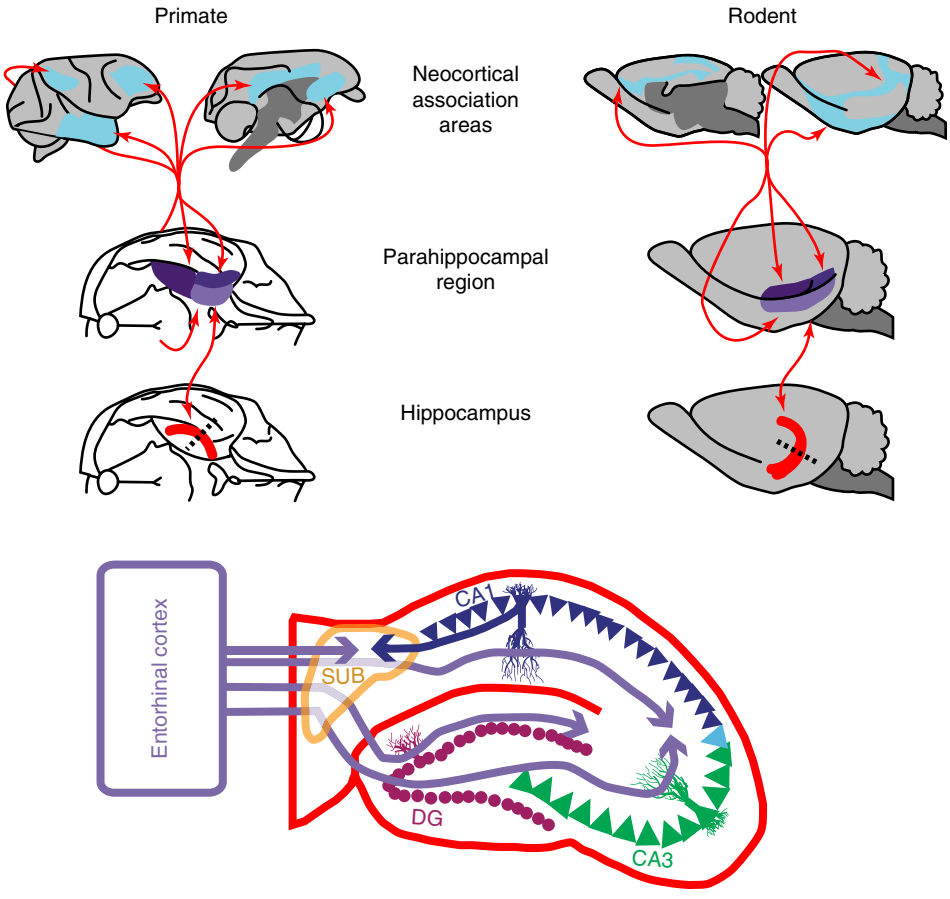


FIGURE 48.2 The anatomy of the hippocampal memory system in monkeys and rats. The hippocampal memory system includes the hippocampus (CA fields, dentate gyrus, and subiculum) and the parahippocampal region, which includes the entorhinal cortex, the perirhinal cortex, and the parahippocampal cortex. The hippocampal memory system also includes midline diencephalic nuclei. Multiple association areas in the cerebral cortex send outputs that converge on cortical areas in the parahippocampal region, which in turn sends its outputs to the hippocampus. The output path involves return projections from the hippocampus to the surrounding parahippocampal region, which in turn projects back to the same cortical association areas. The bottom image shows detailed entorhinal projections on a transverse section of the hippocampus (location of section indicated by dashed lines on the top image). *Top image from Eichenbaum (2001).*

region are directed in turn back to the same areas of the cerebral cortex that were the source of its inputs. Additional structures have also been included as components of this system, including midline diencephalic structures that are connected to the hippocampus via the fornix.

Patient H.M. and Amnesia

The study of the hippocampal memory system, and memory systems in general, began in earnest with a published case report about a man who underwent brain surgery to treat severe epilepsy and who became arguably the most famous patient in neuroscience (Scoville & Milner, 1957; Squire, 2009). This man, known as H.M., began experiencing seizures at the age of 10, and by his late twenties, he had developed a severe seizure disorder that interfered with his ability to live a normal life. Heavy doses of anticonvulsant medications did not stop the seizures from worsening, and the disorder eventually became debilitating. Acting on the idea that dysfunctional brain tissue in the temporal lobes might be causing the seizures, on September 1, 1953, the neurosurgeon William Scoville removed substantial tissue from the medial aspect of H.M.'s temporal lobes bilaterally, including the hippocampus and the parahippocampal cortical region (Fig. 48.3). In one sense, the surgery was a success. The frequency and severity of the seizures were reduced. In another sense, the outcome was tragic. On April 26, 1955, Brenda Milner, a colleague of both Penfield and Hebb, conducted a neuropsychological examination of patient H.M. His profound memory impairment was immediately obvious. H.M. gave the date as March 1953 and his age as 27 (two years younger than his actual age). He performed poorly on memory tests that involved short stories, word lists, pictures, and a wide range of other materials. Remarkably, it was unclear that he even remembered that he had undergone brain surgery. The severity of H.M.'s memory impairment was shocking: he showed almost no capacity for new learning. Yet the reason that H.M.'s case ushered in the modern era of research on memory stemmed from the appreciation that severe memory impairment could occur while other cognitive functions were intact.

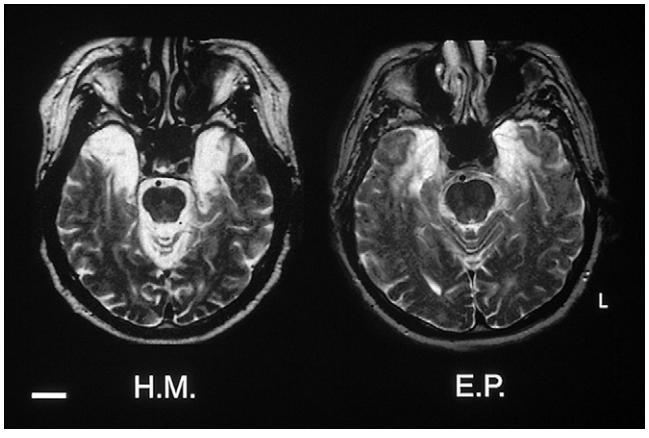


FIGURE 48.3 Magnetic resonance images showing the brains of amnesic patients H.M. and E.P. The images show axial sections through the medial temporal lobes and reveal damaged tissue as a bright signal. H.M.'s damage resulted from surgery, and E.P.'s damage was caused by viral encephalitis. Nevertheless, the resulting lesion was similar for the two patients. Both patients sustained extensive damage to the medial temporal lobes and are profoundly amnesic (from Stefanacci, Buffalo, Schmolck, & Squire, 2000).

First, formal testing identified that H.M.'s IQ score was generally unaffected by the surgery, and a battery of tests found no deficits in perception, abstract thinking, or reasoning ability. These intact abilities indicated that memory could be separated from perception and intelligence. Second, H.M. could hold on to small amounts of information as long as he was actively rehearsing the information. This finding suggested that the ability to maintain information online (now usually referred to as working memory) was distinct from the ability to make a lasting record in the brain (see Chapter 50). Third, H.M.'s childhood memories were relatively intact. This finding suggested that, although the medial temporal lobes might be important for forming new memories, this region was unlikely to be the final storage site for memory. Fourth, H.M. had an intact ability to acquire new motor and perceptual skills (Milner, 1962). As an example, over several days of practice, H.M. gradually improved at tracing the outline of a star when viewing the paper only through a mirror (a task that is initially challenging even for healthy individuals) despite never forming a conscious memory for the testing experience. These results indicated that memories outside the scope of conscious recollection depended on structures outside the medial temporal lobe.

Patient H.M.'s memory impairment is referred to as amnesia, and it includes both an anterograde component—the impaired ability to acquire new information—and a temporally-graded retrograde component—the loss of information acquired soon before the onset of brain damage. Amnesic patients, including H.M., have contributed fundamentally to our current understanding of the hippocampal memory system (see Box 48.1 for a discussion of memory impairments in Alzheimer's Disease). For example, one patient (R.B.) with damage largely restricted to one subfield of the hippocampus (CA1) showed that the hippocampus itself made essential contributions to memory (Zola-Morgan, Squire, & Amaral, 1986), although his memory impairment was more modest than H.M.'s. Perhaps most fundamental has been the repeated demonstration by numerous amnesic patients that the hippocampal memory system supports conscious recollection of facts and events (declarative memory), and that this type of memory is distinct from the type of memory supported by other brain systems, collectively referred to as nondeclarative memory. In most amnesic patients, the deficit in declarative memory encompasses all stimulus modalities and impacts nonverbal expression as well

BOX 48.1

ALZHEIMER'S DISEASE

Alzheimer's Disease (AD) is the most common disorder of memory in humans, afflicting approximately one in eight adults over age 65 (Alzheimer's Association, 2012). The only FDA-approved therapeutics for AD to date are acetylcholinesterase inhibitors and memantine, an NMDA-receptor antagonist, that provide palliative treatment. Since there is no disease-modifying treatment that slows or halts the progression of AD this represents one of the greatest healthcare crises to date. Two characteristic pathological hallmarks of AD are extracellular plaques of aggregated amyloid beta (Aβ) and intracellular neurofibrillary tangles (NFTs) composed of aggregated tau, a microtubule binding protein. Aβ is produced through the enzymatic cleavage (via secretases) of amyloid precursor protein (APP) and in some forms is toxic to neurons. Aβ can disrupt synaptic plasticity, including long-term potentiation. NFTs may compromise intracellular transport and—together with Aβ—kill neurons. Genetic studies from familial cases of AD have strongly pointed towards a causal role for amyloid in aggressive, early-onset forms of AD, with APP or secretase mutations sufficient to cause early-onset AD. However, it is important to realize that familial AD accounts for <1% of all AD cases, while >99% of AD cases are sporadic in nature. Finally, there are a variety of other pathophysiological processes also involved in AD including perturbations in energy metabolism, membrane traffic and signaling, oxidative stress, and inflammation. The precise sequence of events and their relative importance remain to be determined.

Neurodegeneration in AD progresses sequentially through certain brain structures and select subpopulations of vulnerable neurons (Holtzman, Morris, & Goate, 2011). AD almost invariably begins with an insidious, gradual decline in memory, specifically, with problems encoding new memories. Consistently impacted early are those neurons in layer II of the entorhinal cortex, which acts as a gatekeeper of information flowing to and from the hippocampus. Aβ and tau pathologies accumulate in a circuit-dependent manner, deafferenting the entorhinal cortical projections to the hippocampus, and disrupting the molecular mechanisms necessary to form new episodic memories. This results in patients becoming more forgetful and repeating conversations, losing their belongings, and often getting lost driving or otherwise spatially navigating. These initial disease symptoms reflect the increasing dysfunction of the medial temporal lobe memory system. Neuropsychological tests may be used to confirm memory loss and quantify problems of new learning, recall, and recognition memory for verbal and visual material. Additionally, functional MRI studies have demonstrated altered patterns of memory-related activation in the medial temporal lobe of individuals with prodromal AD. While the coding of nascent episodic memories is compromised in early AD, older semantic memories that have previously been consolidated in neocortex are relatively preserved.

As Aβ and tau pathologies become increasingly widespread with disease progression and lead to synaptic failure and neurodegeneration, early episodic memory impairment gives way to progressive deficits in other higher brain functions (Holtzman, Morris, & Goate, 2011). The number of affected neuronal circuits increases dramatically to impact frontal, cingulate, and parietal cortices such that executive function, language, perception, praxis, and visuospatial abilities become impaired. With advanced disease, damage to higher cortices can also disrupt skilled motor actions (praxis) and compromise long-term semantic memory. Beyond widespread cortical pathology, cholinergic basal forebrain (CBF) neurons along with neurons of the locus coeruleus (LC) and raphe nuclei are vulnerable. CBF degeneration contributes to AD symptoms including attention deficits, increased spatial memory decline and further impairment in the coding of new episodic memories. Loss of serotonergic raphe projections likely contributes to behavioral symptoms such as irritability, apathy, depression, mood lability, and aggression. Degeneration of LC neurons translates to noradrenergic transmission loss, which disrupts arousal, vigilance, sleep-wakefulness, working memory, and properly interpreting and responding to autonomic stressors. Additionally, LC degeneration can also perturb microglial function, yielding neurotoxic inflammation.

The pathophysiology of AD allows for several putative therapeutic avenues. At present, our systems-level understanding of AD has yielded pro-cholinergic treatment strategies including acetylcholinesterase inhibitors and NMDA antagonists for transmitter-based symptomatic treatments. Drug candidates are being advanced in clinical trials so that we might obtain the desperately needed therapeutics to slow or halt AD progression. A cellular and molecular understanding of AD has led to amyloid-based approaches including secretase inhibitors, disrupting Aβ aggregation and promoting amyloid clearance. Also under investigation are neurotrophin-based therapies to prevent the neurodegeneration of vulnerable neurons, anti-inflammatories, and other approaches. While much progress has been made in the mechanistic understanding of AD progression, progress in our circuit- and cellular-level understanding of functional deficits and therapeutic benefits *in vivo* has remained particularly challenging. Fully resolving this therapeutic impasse will likely require merging current histology and medicinal

BOX 48.1 (*cont'd*)

chemistry with increased systems-level understanding—through electrophysiology and imaging approaches—of how relevant neural circuits are being modulated pathologically and therapeutically *in vivo* so that our forward progress might be better informed.

Evan P. Lebois and Allan I. Levey

References

Alzheimer's Association. (2012). 2012 Alzheimer's disease facts and figures. Alzheimer's and Dementia. *The Journal of the Alzheimer's Association*, *8*, 131–168.

Holtzman, D. M., Morris, J. C., & Goate, A. M. (2011). Alzheimer's disease: The challenge of the second century. *Science Translational Medicine*, *3*(77), 1–17.

as verbal report whenever the memory task requires the explicit expression of memory, as in tests of free recall or recognition memory.

One profoundly amnesic patient, E.P., has a memory impairment and brain damage very similar to that of patient H.M. (Fig. 48.2). Patient E.P. has shown normal performance on a wide variety of nondeclarative memory tasks, despite the fact that his nearly complete anterograde amnesia has kept him from forming conscious memories of the testing situations themselves. For example, in one study, E.P. was shown eight pairs of objects one at a time, each pair containing a correct object and an incorrect object. He slowly learned to pick up the correct object in each pair and after 18 weeks was able to select the correct object almost every time (Bayley, Frascino, & Squire, 2005). When he was asked how he knew which object to pick up, he pointed to his head and replied, "It's here somehow or another and the hand goes for it." In addition to being unavailable to awareness and being very slowly acquired, E.P.'s memory for the objects appeared to lack the flexibility of that shown by healthy individuals, who had learned the task after only three sessions. When the task was changed by presenting all 16 objects together and by asking E.P. to sort them into correct and incorrect piles, his performance deteriorated precipitously. In contrast, healthy individuals performed nearly perfectly on the sorting task. In another study, over the course of 12 weeks, E.P. came to learn a list of 60 three-word nonsense sentences (e.g., "Speech caused laughter."), such that he eventually was able to respond correctly some of the time when asked to fill in the missing third word (e.g., "Speech caused ???"; Bayley & Squire, 2002). Yet E.P. continually expressed surprise when informed that he had seen the sentences previously. On this task, his memory also appeared more rigid than that of control subjects. When the middle word of each sentence was replaced by a synonym, E.P. was able to answer correctly for only one sentence. By contrast, healthy individuals who learned the sentences after only two weeks had no trouble adapting to the new testing format. These studies with patient E.P. illustrate that memory supported by structures outside the hippocampal memory system is not limited to simple and reflexive behaviors, but can be complex and can include acquisition of new verbal information. These studies also highlight the characteristics of memory supported by the hippocampal memory system. Memories formed via the hippocampal system are typically rapidly acquired, flexible, and available to conscious recollection.

Models of Anterograde Amnesia in Experimental Animals

The memory impairment in human amnesic patients has been closely modeled by the performance of monkeys with similar brain damage (Zola-Morgan & Squire, 1985). One key to establishing a model of amnesia in experimental animals was developing tasks suitable for monkeys that would test declarative memory. Figure 48.4 shows two of these memory tasks. Using one of these tasks, researchers found that monkeys with brain damage similar to H.M.'s performed poorly when they had to retain information about objects for more than a few seconds (Mishkin, 1978). Thus, monkeys with medial temporal damage, like humans with amnesia, have intact immediate memory but perform poorly when the memory demand increases over time. In H.M., the damage included the amygdala, the hippocampus, and the surrounding parahippocampal region. Studies in monkeys with restricted lesions to different structures in the medial temporal lobe have shown that the amygdala is not necessary for declarative memory. In addition, the severity of memory impairment depends on the extent and locus of damage within the medial temporal lobe. Damage limited to the hippocampus, or to its major connections through the fornix, produces only a modest impairment. By contrast, damage that includes the parahippocampal cortices produces severe amnesia. Thus, while the hippocampus is a critical component of the system (Zola et al., 2000), the perirhinal and parahippocampal cortical regions also make major contributions to memory. A focus of current research is to better understand the unique contributions of each of the structures within the hippocampal memory system.

Studies in rodents with damage to the hippocampal memory system have also increased our understanding of the functions of the hippocampal memory system.

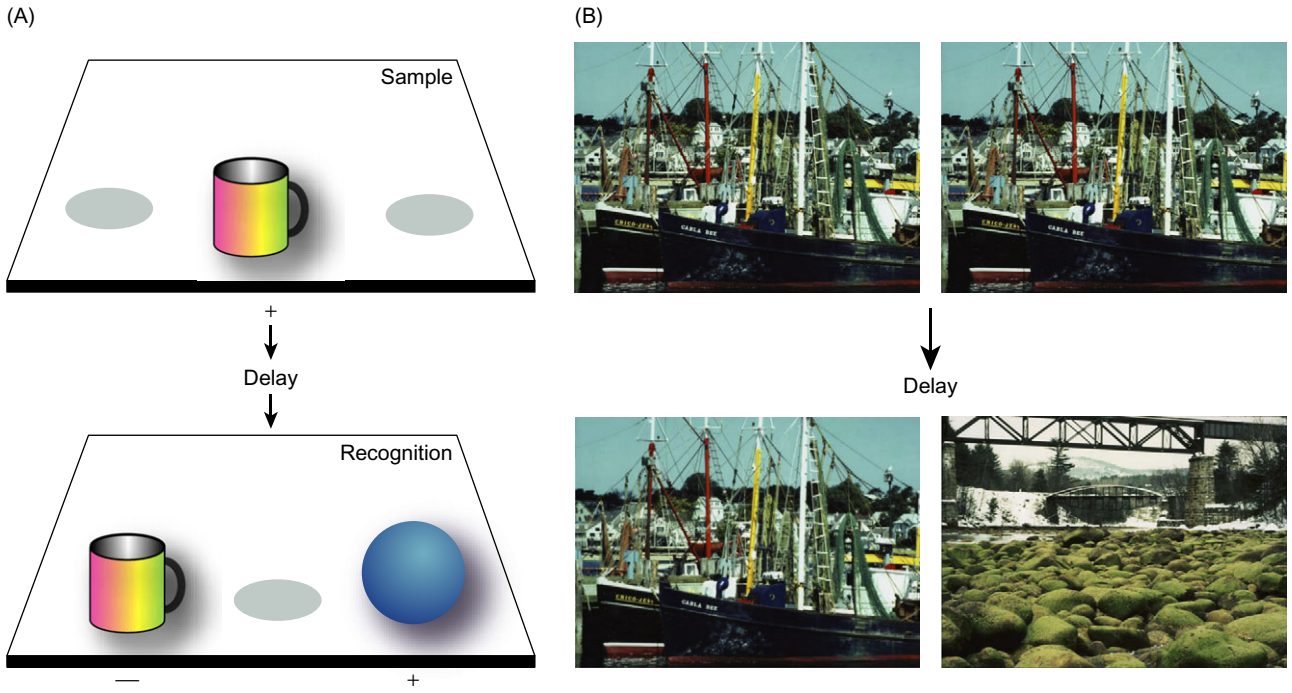


FIGURE 48.4 Recognition memory tasks used for studies of memory in nonhuman primates. Both tasks are easily adapted for human participants as well. (A) The delayed nonmatching-to-sample task using unique objects as stimuli. The subject is initially presented with a single novel object as the sample and must displace the object. This is followed by a variable delay during which the subject cannot see any objects. In the subsequent recognition test, two objects are presented, one of which is the same as the sample and the other of which is novel. Correct performance requires the subject to recognize and avoid the sample object and instead choose the novel one to receive a food reward. (B) Visual paired comparison task. During the sample phase the monkey looks at two identical pictures. In the test phase, one of the sample pictures is represented along with a novel picture. Memory for the repeated picture is inferred by measuring the subject's tendency to look away from the repeated picture and towards the new picture.

Rats with damage to the hippocampus are typically impaired on maze-based tasks when the task depends on memory for space that is capable of being flexibly expressed. One example is place learning in the Morris water maze task (Morris, Garrud, Rawlins, & O'Keefe, 1982; Fig. 48.5). In this task, rats are trained to find a hidden escape platform submerged just below the surface in a pool of cloudy water. The rats are unable to see the platform and must learn the location of the escape site on the basis of cues that are visible on the walls of the room around the pool. Rats with hippocampal damage can learn to locate the escape platform when they start from the same location on every trial. However, their performance is poor when they are asked to demonstrate flexible memory for those spatial relationships by swimming to the remembered location of the submerged platform from a different start point. Consistent with the studies with patient E.P., this pattern of performance reflects a more rigid memory in the absence of a hippocampus.

Other experimental evidence in rats indicates that the hippocampus is also critically involved in the organization and flexible expression of nonspatial memories. For example, in one experiment rats were trained to dig through scented sand to obtain a cereal reward (Bunsey & Eichenbaum, 1996; Fig. 48.6). Initially, both healthy rats and rats with hippocampal lesions learned a set of simple associations between pairs of odors. For example, rats learned to dig in odor B when presented with odor A and to dig in odor C when presented with odor B. Subsequently, the rats were given probe tests to determine the extent to which learned representations supported two forms of flexible memory expression. One of these tests, a test for transitivity, measured the ability to infer an association between two odors that shared a common associate. For example, having learned that odor A is associated with odor B and that odor B is associated with odor C, could they infer that A is associated indirectly with C? The other test, a test for symmetry, measured the ability to recognize associated odors when they were presented in the reverse of their training order. For example, if B is associated with C, is C associated with B? Intact rats showed strong transitivity and symmetry but rats with damage to the hippocampus showed no evidence of transitivity or symmetry. These findings, along with the findings from the water maze and from monkeys, suggest that the hippocampus contributes to rapidly acquired and flexibly expressed spatial and nonspatial memories just as it does in humans.

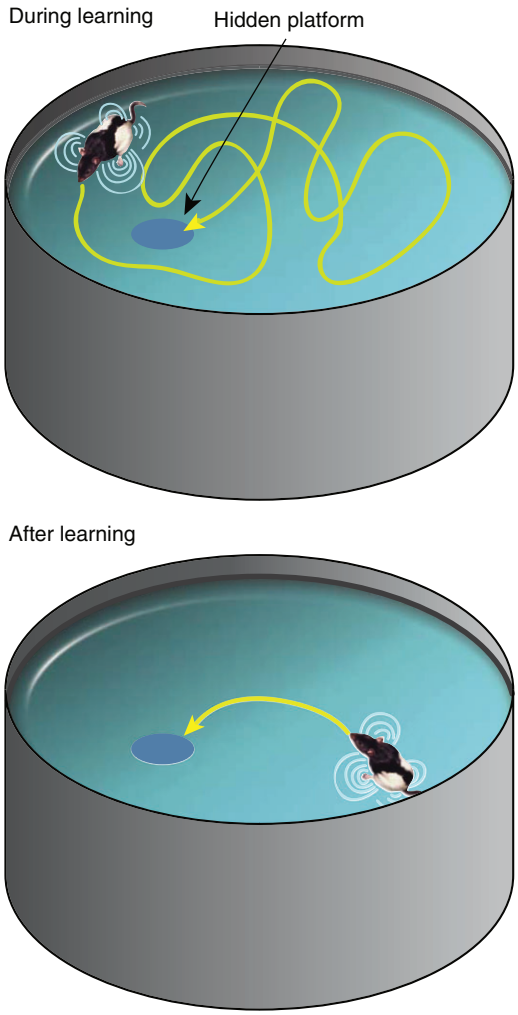


FIGURE 48.5 The Morris water maze task. Early in training, rats search for the submerged platform for extended periods. After training the rat swims directly to the platform.

Neural Activity that Supports Declarative Memory

Studies with humans and experimental animals with damage to the hippocampal memory system have identified the types of memory that depend on the integrity of the hippocampus and parahippocampal region. Complementary studies involving neurophysiological recording and imaging of neural activity in intact medial temporal lobe structures have demonstrated how the hippocampal memory system ordinarily functions in support of declarative memory. In particular, noninvasive functional brain imaging, including functional magnetic resonance imaging (fMRI), has become a key tool for understanding how the hippocampal memory system operates in the intact human brain. In one of the earliest studies, increased activity was observed in the hippocampus and parahippocampal region when participants viewed previously unseen photographs as compared to when participants viewed photographs that they had already seen on multiple occasions (Stern et al., 1996). Subsequent studies found that activity in the hippocampal memory system is often related to the success of memory encoding or retrieval. For example, in one study, participants saw a list of words one at a time while being scanned (Wagner et al., 1998). After the scanning session, participants were given a test to assess which words were remembered and which words were forgotten. The results indicated that activity in the hippocampal memory system during encoding (in this case, in the parahippocampal region) was greater for words that were subsequently remembered compared to words that were forgotten. More recent studies suggest that, in humans, the different cortical areas of the parahippocampal region may play distinct roles in memory, with the perirhinal cortex showing increased activity during the encoding of specific objects and the parahippocampal cortex showing increased activity during the encoding of locations (Buffalo, Bellgowan, & Martin, 2006; Litman, Awipi, & Davachi, 2009; Pihlajamaki et al., 2004). These findings suggest that potential functional differences among medial temporal lobe regions may be linked to distinct neuroanatomical connectivity from nonspatial and spatial processing streams to these regions (Suzuki & Amaral, 1994; Witter et al., 2000).

While neuroimaging studies allow for an identification of neural activity averaged across a few millimeters of brain tissue, intracranial neurophysiological recording techniques allow for an identification of neural activity at the level of individual neurons. Studies in which electrodes were surgically implanted in the brains of monkeys, rats, and even humans (in cases of patients preparing for neurosurgery) have shown that the spiking activity of individual hippocampal neurons corresponds with memory performance on numerous tasks. For example, evidence of novelty and familiarity signals was found in a study of monkeys performing a simple visual recognition memory task (Jutras & Buffalo, 2010), similar to the paired comparison task described above (Fig. 48.4B). Across trials, many hippocampal neurons showed stronger responses to novel stimuli than to repeated stimuli (Fig. 48.7). Further, the magnitude of these firing rate modulations was positively correlated with memory performance. Larger changes in firing rate between novel and repeated stimuli were significantly correlated with better recognition memory performance. Results from human recordings studies also suggest that hippocampal neurons can represent stimulus novelty or familiarity, independent of the specific stimulus (Rutishauser, Mamelak, & Schuman, 2006). Importantly, these results provide evidence linking hippocampal activity with recognition memory performance.

Other studies have shown that hippocampal neurons can demonstrate stimulus-specificity. For example, neurons in the human hippocampus have been shown to fire preferentially for specific categories of images—for

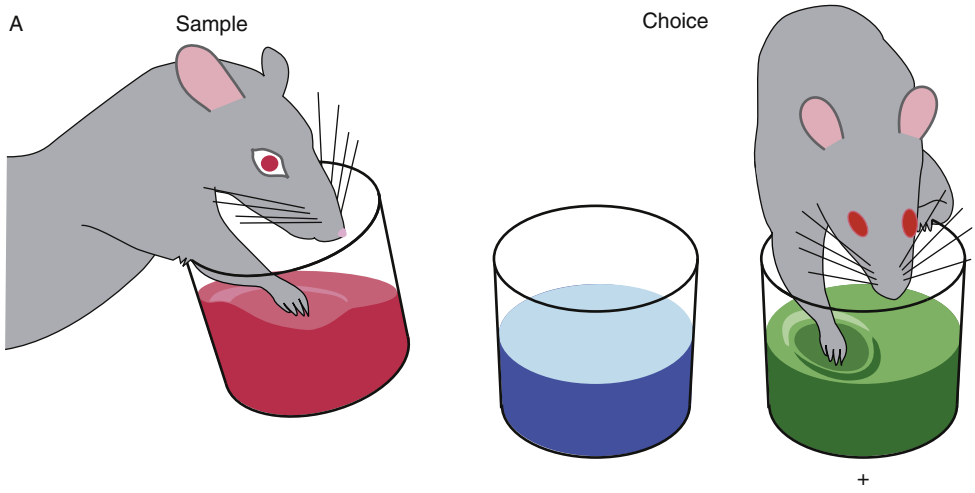

A
Sample
Choice
+
B
Odor paired associates
Training set 1: AB & XY
Sample A X
Choice B+ vs Y B vs Y+
Test for Transitivity: AC & XZ
Sample A X
Choice C vs Z C vs Z
Training set 2: BC & YZ
Sample B Y
Choice C+ vs Z C vs Z+
Test for Symmetry: CB & ZY
Sample C Z
Choice B vs Y B vs Y

FIGURE 48.6 Associative transitivity and symmetry in paired associate learning. (A) On each training trial, one of two odors is presented as the sample. On the subsequent choice trial the animal must select the assigned associate, indicated by a "+". (B) Outline of odor pairings used in training on two sequential sets of paired associates, plus stimuli used in tests for transitivity (C for A; Z for X) and symmetry (B for C; Y for Z). *From Bunsey and Eichenbaum (1996).*

example, faces versus natural scenes (Kreiman, Koch, Fried, 2000) and when the participant viewed or recalled specific video clips (Gelbard-Sagiv, Mukamel, Harel, Malach, & Fried, 2008). Hippocampal neurons in monkeys have been shown to fire preferentially when specific objects (Naya & Suzuki, 2011) or visual scenes were presented to the monkey, and the firing rates of these neurons showed modulations in activity that correlated with the monkeys' ability to learn to move their eyes to a particular place in the scene to receive a reward (Wirth et al., 2003). Studies in rats have also shown that hippocampal neurons sometimes fire preferentially for specific olfactory stimuli as well as combinations of these stimuli and the places where they occur, suggesting that hippocampal neurons represent important stimuli in the context in which they are remembered (Wood, Dudchenko, & Eichenbaum, 1999).

In addition to this activity related to memory for items, neurons with spatial correlates have been identified in the hippocampal memory system of several species of mammals, including humans (O'Keefe & Dostrovsky, 1971; Derdikman & Moser, 2010; also see Chapter 45). In particular, many neurons in the rat hippocampus, referred to as "place cells," have clearly defined spatial receptive fields, or "place fields" (Fig. 48.8A). Activity of these cells reflects the constellation of auditory, olfactory, tactile, and visual cues that define the external world as well as the self-motion cues that track the rat's path through that space. Yet in addition to these clear spatial correlates, activity of these cells also reflects factors such as goals and expectations, and a common view is that the type of spatial memory reflected by place cells in rats shares many commonalities with examples of memory in humans such as recall. Further, cells with location or direction correlates have been found in several regions interconnected with the hippocampus, leading to an understanding of how mental maps might be constructed in the brain. An important advancement in that goal was the identification of "grid cells" in a region of the medial entorhinal cortex, an area reciprocally connected with the hippocampus (Hafting, Fyhn, Molden, Moser, & Moser, 2005). The activity of one of these cells tiles a recording chamber in

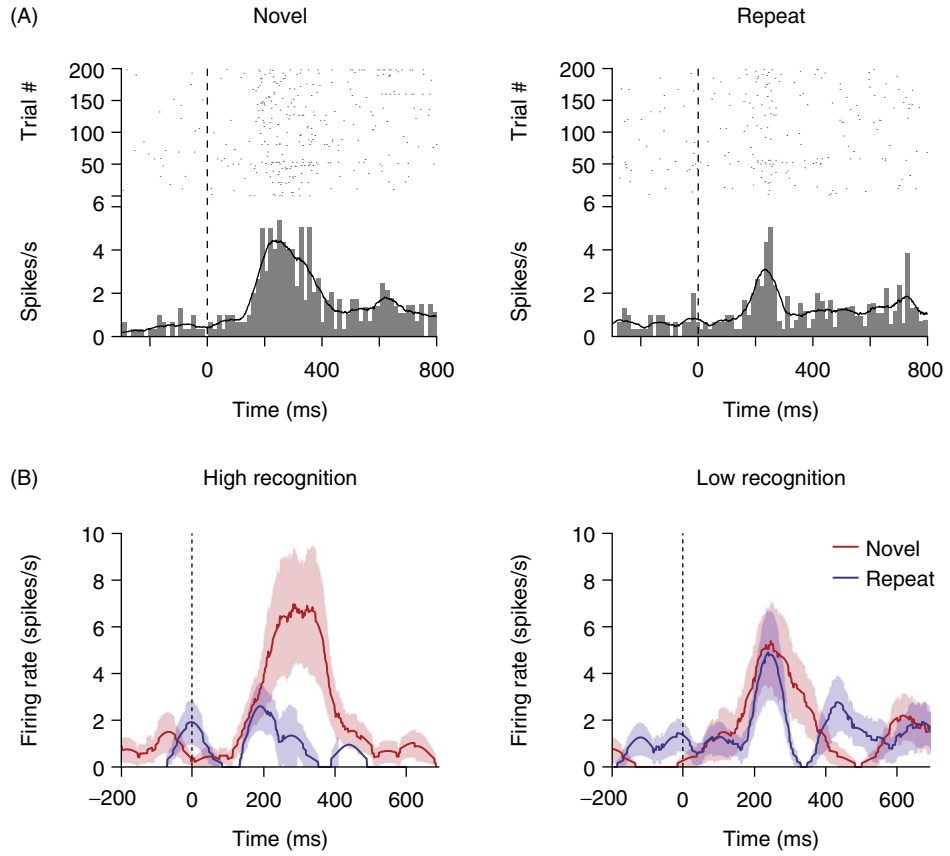


FIGURE 48.7 Recognition memory signals in the monkey hippocampus. (A) As monkeys performed a variant of the Visual Paired Comparison task, some hippocampal neurons showed a change in firing rate that reflected stimulus novelty. This neuron fired more when the pictures were novel (Left panel) compared to when they were repeated (Right panel). (B) Many neurons showed a greater modulation in firing rate by novelty for pictures that were better remembered (Left panel) compared to those that were not well remembered (Right panel). *From Jutras and Buffalo (2010).*

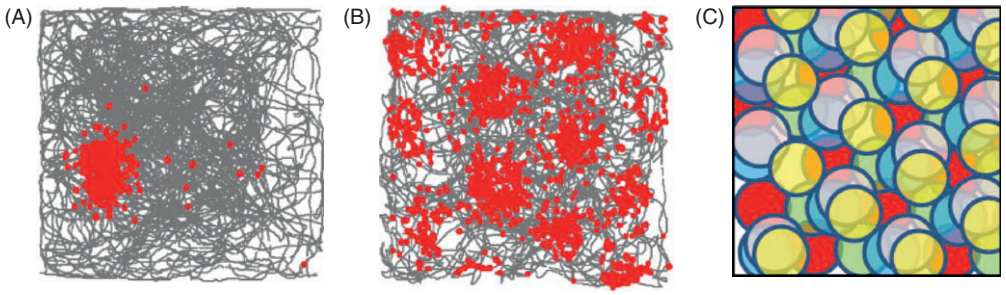


FIGURE 48.8 An example of a place cell (A) and a grid cell (B) recorded from a rat hippocampal and entorhinal neuron, respectively. The gray lines show the path of a rat as it explored a square recording chamber, and the red dots show the locations along that path at which an example hippocampal (A) or entorhinal (B) neuron fired action potentials. The right panel (C) illustrates activity from multiple grid cells, and one idea is that place fields might relate to points of overlap from a collection of grid fields. *From Derdikman and Moser (2010).*

a geometrically regular grid pattern, and the overlapping receptive fields of the population of grid cells in a rat's brain are thought to be an essential component of the rat's mental map (Fig. 48.8B). An important question is how the circuitry of the hippocampus and entorhinal cortex (and other connected areas) enables the transformation of spatial information from a grid field to a place field (and back again). A further question is how nonspatial information is incorporated into these spatial representations to support remembering, for example, which item was encountered in a given location. These questions highlight the distributed nature of memory representations, and the answers could be relevant to general information processing principles in the mammalian hippocampal memory system.

In addition to recording activity of individual neurons, electrode recordings can measure the summed activity of many neurons in the local field potential at a particular recording site (measured either noninvasively from the scalp or with electrodes implanted on the surface or deep in the brain). Using this approach, studies have demonstrated memory-related changes in coordinated network activity that is reflected through oscillations, the synchronized and rhythmic activity of many neurons. Neuronal ensembles often synchronize their activity at a particular frequency band, and oscillatory activity during memory tasks has been observed at theta-band (4 to 12 Hz) and at gamma-band (30–90 Hz) frequencies.

Gamma-band oscillations have been identified in neural recordings taken from patients who volunteered to participate while they were awaiting neurosurgery. In one experiment, patients were asked to study a list of words and subsequently were asked to freely recall as many words as possible. Interestingly, the power of gamma-band oscillations was greater in the hippocampus and parahippocampal regions during the encoding of subsequently recalled words compared to unrecalled words (Sederberg et al., 2007). Using a similar task, others showed that interaction between medial temporal lobe structures might be important for memory formation. In this study, Fell et al. (2001) demonstrated that synchronized activity between the hippocampus and parahippocampal region in the gamma-frequency band was higher during the encoding of subsequently recalled words compared to unrecalled words. Recently, the relationship between gamma-band synchronization and memory formation was extended to single hippocampal neurons. The activity of hippocampal neurons was recorded while monkeys performed a preferential looking task, similar to the visual paired comparison task. Hippocampal neurons showed enhanced gamma-band synchronization during the encoding of pictures that the monkeys later recognized well (Jutras, Fries, & Buffalo, 2009). The time-course of this enhancement was extremely similar in monkeys and humans, using a different behavioral paradigm, which suggests that gamma-band synchronization may reflect a basic mechanism for the neuronal interactions that are critical for successful memory encoding (Fig. 48.9).

Slower, theta-band oscillations are a ubiquitous feature of hippocampal activity in rodents, especially when the animals are actively exploring an environment, and theta-band oscillations have also been described in cats, bats, and even humans. Modulations in theta-band oscillations have also been linked to memory formation (Fuentemilla, Penny, Cashdollar, Bunzeck, & Duzel, 2010; Rutishauser, Mamelak, & Schuman, 2010). For example, Rutishauser et al. (2010) demonstrated that an increase in the tendency of hippocampal neurons to emit action potentials at a consistent phase of the theta-band oscillation was associated with better memory, indicated by higher retrieval confidence in human patients.

Oscillations play an important role in synchronizing activity among neurons within a region or between regions, and recent studies suggest that theta-band and gamma-band oscillations interact to coordinate the flow of information in the hippocampal memory system. For example, gamma-band oscillations have been observed to ride on top of theta-band oscillations, with a single theta wave containing several gamma waves. A recent study in rats found that the troughs of theta waves in CA1 tended

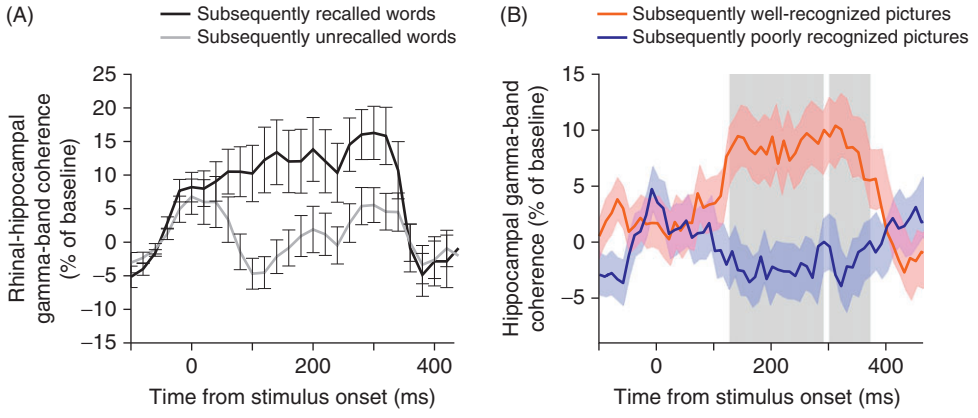


FIGURE 48.9 Gamma-band synchronization in the medial temporal lobe during memory encoding is associated with the degree of subsequent recognition. (A) Gamma-band phase synchronization (coherence) between the human hippocampus and the rhinal cortex during word study, as a function of time from stimulus onset. Coherence was significantly higher during the encoding of words that are subsequently recalled (black) than for words that were not later recalled (gray). Error bars indicate SEM. *Adapted by permission from Macmillan Publishers Ltd. from Fell et al., 2001.* (B) Gamma-band spike-field coherence in the monkey hippocampus during the encoding of pictures, as a function of time from stimulus onset. Coherence was significantly higher for stimuli to which monkeys subsequently showed a high degree of recognition (red) than for stimuli which were not well recognized (blue). Red and blue shaded areas represent SEM. Gray shaded area represents time points at which gamma-band coherence was significantly different for the two conditions ($p < 0.01$). *Modified from Jutras et al. (2009).*

to carry a type of fast (around 65–150 Hz) gamma oscillation also prominent in entorhinal cortex, whereas the downward slope following the peaks of the theta waves tended to carry slow (25–50 Hz) gamma oscillation also prominent in CA3 (Colgin et al., 2009). These results are consistent with the idea that inputs to CA1 from CA3 and entorhinal cortex (Fig. 48.3) are precisely timed by oscillatory activity in the hippocampal memory system in order to optimize memory performance (Hasselmo, Bodelon, & Wyble, 2002; Manns, Zilli, Ong, Hasselmo, & Eichenbaum, 2007). Indeed, action potentials occurring within a single gamma wave that is nested in a theta wave will occur within about 10 ms of each other, a time window that is optimal for synaptic plasticity (Bi & Poo, 1998). Thus, oscillations likely provide the brain with the means to coordinate the activity of large cell assemblies while also optimizing the plasticity within individual junctures of that network.

Transgenic Animals and the Dissection of Hippocampal Circuitry

Experimental animals (typically mice) with modified genes have become another important avenue for exploring the functional circuitry of the hippocampal memory system. Researchers have combined techniques in cell biology with selective breeding to produce transgenic animals that produce too much or too little of targeted proteins in specific parts of the brain. In addition, the altered protein production can be brought under control of drugs that the animals ingest, allowing the timing of the expression of the genetic manipulation to be controlled in addition to the location in the brain. One very fruitful approach to understanding how the hippocampus contributes to declarative memory has been to test the memory of mice that lack a protein important for synaptic plasticity or synaptic transmission in a specific type of cell (e.g., pyramidal cells in hippocampal area CA3). In this way, one can ask how impairing memory at a particular juncture at the synapse level will relate to memory impairments at the systems level. For example, a recent study used sophisticated techniques to produce a strain of mice in which synaptic transmission between CA3 and CA1 hippocampal neurons could be temporarily blocked (Nakashiba, Young, McHugh, Buhl, & Tonegawa, 2008). The combination of memory testing (e.g., in the Morris Water Maze) and neural recording in downstream CA1 pyramidal cells in these mice suggested that the entorhinal cortex input to area CA1 was sufficient for normal acquisition of spatial memory over several trials but that the CA3 input to CA1 was important for single-trial acquisition of spatial memory.

The combination of genetic approaches with animal models of amnesia and neuronal recording has started to sketch some of the roles of the structures within the hippocampus. Two important aspects of memory that have been proposed to differentially involve the dentate gyrus, CA1, and CA3 subregions of the hippocampus are termed "pattern separation" and "pattern completion." Pattern separation refers to the process of distinguishing between two similar inputs and storing them as separate representations. By comparison, pattern completion refers to the process of filling-in partial or incomplete sensory input based on stored representations. There is accumulating experimental evidence and computational theory to support the idea that the dentate gyrus input to CA3 is a particularly important pathway for pattern separation, while CA3 neurons combine input from the dentate gyrus, the entorhinal cortex, and their own recurrent collaterals and project to CA1 in the process of pattern completion (Rolls, 2010; Yassa & Stark, 2011).

Retrograde Amnesia and Consolidation

So far, this chapter has focused primarily on the structures and neural signals that support the formation of new memories. However, patients with amnesia due to medial temporal lobe damage suffer not only a deficit in learning new material (anterograde amnesia), but also loss of memories that were acquired before the brain damage (retrograde amnesia). Importantly, the retrograde deficit is time-limited, and material acquired shortly before the damage is affected most severely, whereas items learned earlier in life are relatively spared. For example, the profoundly amnesic patient E.P. (Fig. 48.2) has almost no capacity for new declarative learning but shows intact memories from his childhood. In one study, E.P. was asked to think back to his childhood neighborhood and describe routes he would take from one location to another (e.g., his grade school to the town movie theater; Teng & Squire, 1999). In some instances, he was asked to imagine that a main street was blocked and to describe an alternate route. E.P. performed as well as a group of healthy individuals who had lived in the neighborhood at the same time and, like E.P. had moved from the area as young adults, long before the study. These data indicated that his memory for the spatial layout of his childhood neighborhood had escaped the damage to his medial temporal lobes. Consistent with his severe anterograde amnesia described earlier in this chapter, E.P. has been unable to learn the layout of the neighborhood he moved to subsequent to the onset of amnesia.

Retrograde amnesia is another aspect of the amnesic syndrome that has been studied in experimental animals. For example, studies have shown that simple object discriminations that were learned by monkeys shortly before medial temporal lobe damage are poorly retained, but discriminations learned remotely are spared. This pattern of memory impairment has been replicated in a variety of tasks in humans, monkeys, rabbits, rats, and mice and is thought to reflect a process of memory consolidation (Squire, Clark, & Knowlton, 2001).

The term *consolidation* has been used to characterize two kinds of brain events that affect the stability of memory after learning. One event involves the fixation of plasticity within synapses over a period of minutes or hours through a sequence of protein synthesis and morphological changes at synapses (see Chapter 47). The other event involves a reorganization of memories, which occurs over weeks to years following new learning. This prolonged consolidation occurs in the hippocampal memory system and is thought to involve interactions between the hippocampus, parahippocampal region, and the cerebral cortex. Several models have been proposed to account for how the hippocampus might interact with other brain regions over a prolonged period in memory consolidation (Alvarez & Squire, 1994; McClelland, McNaughton, & O'Reilly, 1995). These models assume that widespread areas of the neocortex contain the details of the information that is to be remembered and that areas in the hippocampal memory system support the capacity to retrieve the memory during the period shortly after learning. Over time, areas in the hippocampus are thought to reactivate the cortical representations through repetition, rehearsal, or spontaneous activity. The reactivation is then thought to induce plasticity in cortical-cortical connections, and these connections are viewed as a possible network structure for the permanent storage and organization of the memory. In support of this idea, it was recently demonstrated that activity in the neocortex could predict the recall of specific memories. Polyn, Natu, Cohen, and Norman (2005) asked human subjects to first study three lists of items including photographs of famous faces, famous locations, and common objects. Subsequently, the subjects performed a free-recall test where they were asked to recall as many items from the study lists as possible. Functional MRI was used to examine neural activity during both the study episode and the free-recall period, and the data revealed patterns of activity immediately prior to the recall of specific items that were similar to activity observed as the subjects studied that category of items (e.g., faces vs. locations or objects).

Recent studies using molecular markers of neuronal activity and synaptic change have also reported evidence consistent with this idea. For example, one study of maze learning in mice observed that these markers were found in high numbers in the hippocampus one day after learning but their numbers were much lower 30 days later (Maviel, Durkin, Menzaghi, & Bontempi, 2004). By contrast, the same markers were found at low levels in the cerebral cortex initially after learning but were high in various cortical areas 30 days after learning; cortical areas included the prefrontal, anterior cingulate, and retrosplenial cortices. These results suggest that plasticity in the hippocampus is initially important for memory formation, but over time, plasticity in neocortical areas becomes more important.

There has long been interest in the role of sleep and dreaming in memory, and it has been suggested that processes might occur during sleep that promote the consolidation of hippocampus-dependent memories. Although the benefit of sleep for nonconscious learning is generally accepted, the status of hippocampus-dependent memory is less certain (Stickgold, 2005). Nevertheless, there are several results that indicate that processes during sleep might contribute to consolidation of hippocampus-dependent memory. For example, specific patterns of activity observed in the hippocampus while rats were awake were observed to "replay" during both REM sleep and slow wave sleep (Sutherland & McNaughton, 2000). Later studies linked hippocampal replay to sharp wave ripples (SWRs). SWRs are prominent during sleep and are recorded in the CA3 and CA1 regions as a transient (~100 ms) high-frequency oscillation (100–250 Hz). During SWRs, activity propagates from CA3 to CA1, and then out to the neocortex. This hippocampal-cortical interaction may be important for consolidation, and studies suggest that reactivation of memories during sleep may participate in the process of consolidation. SWRs occur also during the waking state, and hippocampal replay has been recently observed during SWRs as rats explored a novel environment. Interestingly, awake replay is more prevalent and temporally precise when rats are learning a novel environment, compared to when they are placed in a familiar environment (Diba and Buzsaki, 2007; Foster and Wilson, 2006). SWRs occur most often during slow-wave sleep, and another study showed that induction of slow-wave-like oscillations in the brains of human participants while they slept led to better retention in that less forgetting occurred across the interval (Marshall, Helgadottir, Molle, & Born, 2006). Thus, there is accumulating evidence regarding the role of sleep in the consolidation of hippocampus-dependent memory, but further work is needed to fully understand this process.

Summary of Hippocampus

Amnesia associated with damage to the hippocampal memory system in humans is characterized by an inability to form and consciously recollect memories of facts and events. Studies with animal models of amnesia have furthered our understanding of the function of the hippocampus and the parahippocampal region. Neurophysiological and neuroimaging studies have provided insight into the neural activity that may underlie declarative memory formation. Taken together, data obtained from a variety of techniques have indicated the importance of interactions within the hippocampal memory system as well as between the hippocampal memory system and the cerebral cortex in the formation and consolidation of declarative memories.

Striatum

Anatomy

The striatum is an important component of the basal ganglia (see Chapter 30), and its position within a distributed cortico-basal ganglia network is central to its function as a hub of a memory system. Sensorimotor, associative, and limbic cortical areas project to specific regions of the striatum (dorsolateral, dorsomedial, and ventral regions, respectively), and outputs of these regions loop through the pallidum to gain access to brainstem structures and back to cortical areas via the thalamus (Fig. 48.10). In addition, the striatum receives modulatory inputs from dopaminergic cells in the midbrain. Thus, the striatum receives information relevant to actions, external stimuli, and internal states (including expectations) and can influence behavior via brainstem pathways or by updating processing in cortical areas. See Chapters 30 and 41 for additional consideration of the role of the striatum in motor control and in reward and motivation, respectively.

Identifying Memory Supported by the Striatum

Early efforts to understand the role of the striatum in memory often sought to distinguish its contributions from those of the hippocampal memory system. For example, one study compared performance on a simple T-shaped maze between rats with pharmacological inactivation of the hippocampus or dorsal striatum (Packard & McGaugh, 1996). During training, neither structure was inactivated, and all rats learned to run to a consistent choice arm (e.g., left for some rats) for a food reward when started at the base of the T-maze. During testing, the T-maze was rotated 180 degrees, and the question was whether a rat would continue to make left turns at the choice point (if it had been trained to run to the left arm), indicating memory resembling a habitual response ("response" memory), or would now make right turns, indicating a more flexible memory for the location of the food reward relative to the surrounding room ("place" memory; Fig. 48.11). The finding was that, after a week of training, most rats showed evidence of using place memory. Furthermore, inactivating the striatum during the test had little effect on performance, but inactivating the hippocampus led some rats to shift to a strategy that suggested the rats were compensating with response memory. Following the initial test, the T-maze was rotated back to its original position, and all rats received another week of training during which neither structure was inactivated. After this second week of training, the T-maze was again rotated 180 degrees, and the results on the second test differed substantially from the first. Most rats now showed evidence of relying on the response memory, and inactivation of the hippocampus had little effect on performance. Instead, inactivation of the striatum led many rats to alter their performance and again resort to place memory. These results suggested that, after initial training, performance normally relied on a memory for the location of the food that was flexible enough to be accessed despite rotation of the maze and that this type of place memory depended on the hippocampus. However, after more extended training, performance appeared to come under influence of a habitual response memory that depended on the striatum.

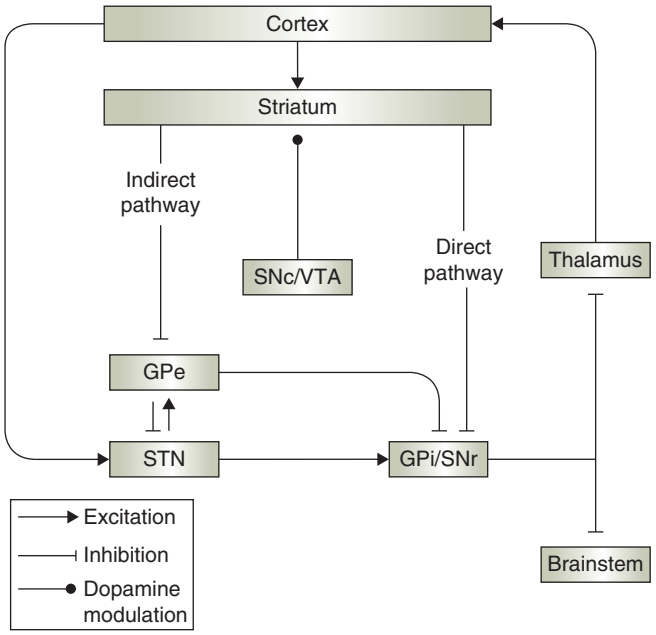


FIGURE 48.10 Anatomy of the striatum and the rest of the basal ganglia. STN, subthalamic nucleus; GPe, external globus pallidus; GPi, internal globus pallidus; SNr, substantia nigra pars reticulata; SNc, substantia nigra pars compacta; VTA, ventral tegmental area. *From Yin and Knowlton (2006).*

These early results with rats were paralleled by a study that compared memory between amnesic patients with damage to the extended hippocampal memory system and patients with Parkinson's disease (Knowlton, Mangels, & Squire, 1996), a disease that disrupts striatal function through degeneration of dopaminergic inputs from the substantia nigra. Both groups of patients and a group of healthy individuals were tested on a task in which they tried to guess on each trial whether the stimulus cards presented to them foreshadowed rain or sunshine (Fig. 48.12). The cards predicted rain or sunshine only in a probabilistic manner, and healthy individuals improved their performance from 50% (guessing) to about 70% only after 50 trials. Thus, the predictive value of the cards did not appear to be easily memorized, and performance instead appeared to rely on slowly acquired trial-and-error learning. Patient's with Parkinson's disease showed little improvement over the 50 trials, but amnesic patients performed about as well as the healthy individuals. In contrast, when the participants were subsequently asked questions about the nature of the task and the kinds of materials they

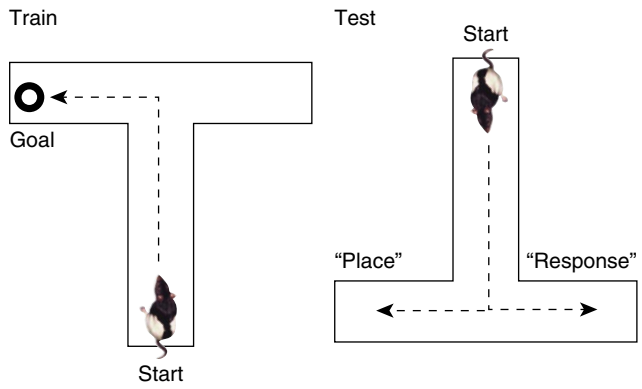


FIGURE 48.11 "Place" versus "response" learning. The rat is trained initially to turn left in order to obtain a reward at a particular location. In a subsequent test, the maze is rotated and the rat is allowed to select whether it will perform the same left turning "response" or remember the "place" of the previous reward.

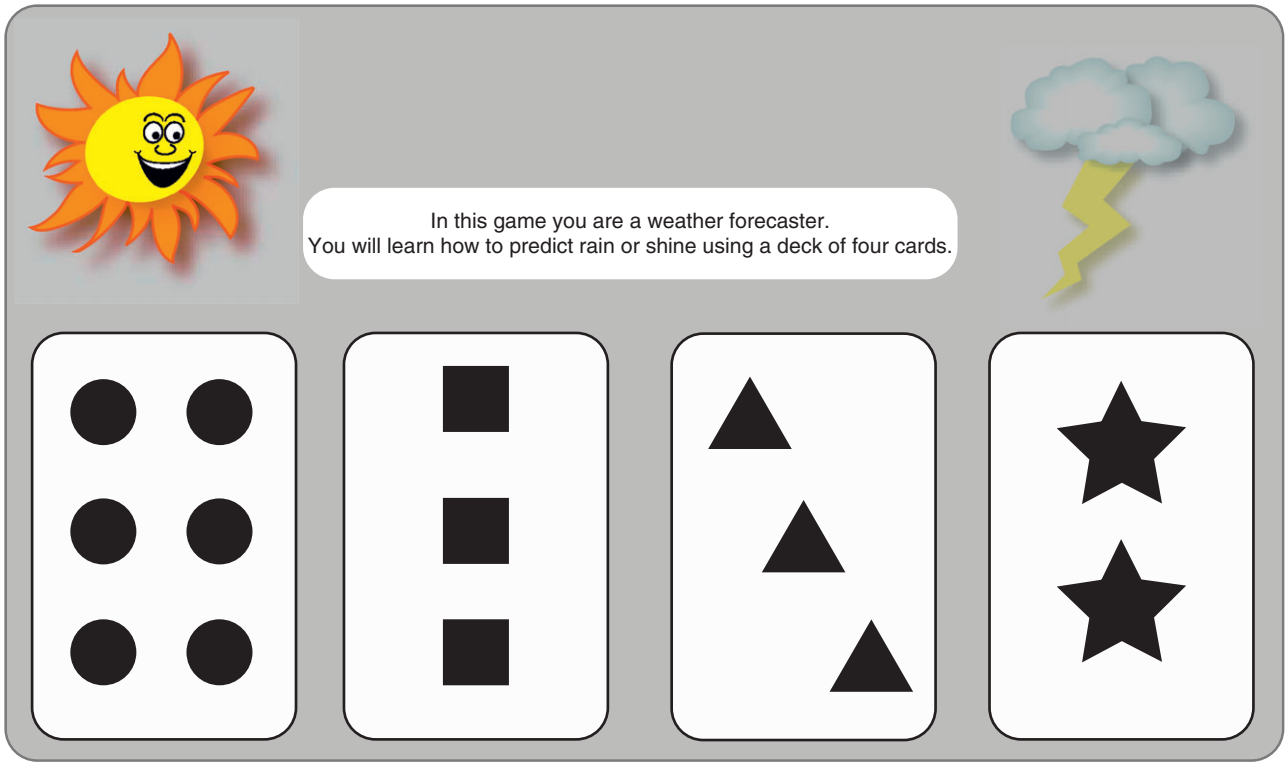


FIGURE 48.12 Weather prediction task. View of the computer screen presented to subjects showing all four stimulus cards and the "sun" or "rain" response choices. *From Knowlton et al. (1996).*

had encountered, the patients with Parkinson's disease performed as well as controls, and the amnesic patients showed very poor memory. Thus, results with humans agreed with results from experimental animals in suggesting that the striatum was part of a memory system that was distinct from the hippocampal memory system.

Memory for Actions, Outcomes, and Habits

These initial studies suggested that the striatal memory system accumulated trial-and-error experiences as inclinations to make certain responses in the presence of particular stimuli and that these inclinations differed from the record of individual experiences supported by the hippocampus. Nevertheless, these studies also suggested that the memory supported by the striatum was too complex to be described fully as simple stimulus-response learning. Indeed, more recent research has investigated how the striatal memory system incorporates outcomes of individual actions into the stimulus-response association and how its initial participation in evaluation and decision making might give way to a role in habit-like responding with extended training.

Studies in which the activity of midbrain dopamine neurons was recorded have provided some clues about how consequences of actions could be incorporated into memory supported by the striatum (Schultz, 2007). These studies have indicated that these midbrain dopamine neurons increase their firing in response not only to rewards but also to secondary cues that predict the delivery of those rewards. In addition, these neurons also appear to signal a mismatch between expected outcome and actual outcome by, for example, an exaggerated response when an unexpected reward was delivered. Thus, the projections into the striatum from these midbrain regions could help to update inclinations or disinclinations to engage a particular response in the presence of particular stimuli based on whether the previous response was rewarded with a favorable outcome.

In addition, studies that have recorded activity in the striatum in both humans and experimental animals have indicated that the striatum likely incorporates these anticipated outcome signals (Graybiel, 2008). For example, a study in which neurons were recorded from the striatum as rats performed a more elaborate version of the T-maze task found that both the dorsal and ventral striatum were engaged at moments when a turning response was to be made but that the contributions of these regions seemed to differ according to whether it was early or late in training (van der Meer, Johnson, Schmitzer-Torbert, & Redish, 2010). In particular, some neurons from the ventral striatum showed increased activity at one of two possible reward locations (i.e., left or right). Some of these reward-related neurons also showed activity early on in training when the rat paused at a point on the maze when the rat needed to choose to turn left or right. Thus, early on in training, the ventral striatum appeared to gain access to prior favorable outcomes in selecting an action to perform in the same scenario again. In comparison, neurons in dorsal striatum showed increased activity unrelated to reward at turns on the maze and did so with more spatial accuracy later on in sessions. These results are consistent with the idea that striatal learning typically progresses from a probing action selection to a habitual response and that this deemphasizing of the outcome parallels a reorganization from a network centered on the ventral striatum to a network more centered in dorsal striatum. This idea has been extended to habits of many kinds, including drug addiction and other compulsive behaviors (see Chapter 41).

Summary of Striatum

The striatum is a key structure in a distributed cortico-basal ganglia memory system that supports the acquisition of inclinations to make specific responses in particular situations. The striatum contributes to an initial evaluative action-outcome learning and to a habitual stimulus-response learning after extended training. Recent results suggest that the ventral striatum might be particularly important for the former and that the dorsal striatum might be particularly important for the latter.

Cerebellum

Anatomy

The cerebellum consists of the cerebellar cortex, which has been divided into lobes and further into lobules, and several deep cerebellar nuclei (see Chapter 31 for more detailed anatomy). The cerebellum receives direct input from the spinal cord and brainstem as well as indirect input from a variety of sensory and motor areas in the cerebral cortex via pontine nuclei in the brainstem. The cerebellum projects directly to the spinal cord, brainstem, hypothalamus, and thalamus, and the thalamic targets in turn project to various motor and nonmotor areas in the cortex, particularly in the frontal lobes (Middleton & Strick, 1998). The internal circuitry of the cerebellum has been described as modular, with a local wiring diagram that is repeated again and again. This modularity raises the possibility that, although different regions of the cerebellum receive different input, each region might perform a similar set of operations on that information before passing the results back out to the rest of the nervous system (Boyden, Katoh, & Raymond, 2004).

Eyeblink Classical Conditioning as a Prototype of Cerebellum-Dependent Memory

The cerebellum is important for a range of sensorimotor functions, including adaptations of the vestibulo-occular response and coordinated motor skill learning. The cerebellum has also been associated with nonmotor learning and other examples of cognition and behavior. The present section focuses on the best-understood example of cerebellar-dependent learning, eyeblink classical conditioning, as a framework for highlighting the role of the cerebellum in memory based on the idea that the modular circuitry of the cerebellum offers a generalized capacity for processing of specific inputs.

In standard "delay" eyeblink classical conditioning, a conditioned stimulus (CS; e.g., a brief tone) is initiated immediately (typically a second or less) prior to the delivery of an unconditioned stimulus (US; e.g., a brief puff of air to the eye), and both stimuli then terminate at the same time. The US triggers a reflexive, unconditioned eyeblink (the UR), and after many pairings of CS and US, the CS eventually elicits an eyeblink as a conditioned response timed such that the blink is most robust at the US onset. Remarkably, rabbits with no cerebral cortex, basal ganglia, limbic system, thalamus, or hypothalamus showed normal retention of the conditioned eyeblink response (Mauk & Thompson, 1987), indicating that no forebrain structures

were an essential part of the neural circuitry supporting standard eyeblink conditioning.

Numerous studies by Thompson and colleagues have traced the paths to the cerebellum taken by incoming sensory information about the CS and the US and have found that these paths converge on one particular cerebellar nucleus, the interpositus nucleus, and that outputs of this nucleus have a clear route to producing conditioned eyeblink responses (Fig. 48.13; Thompson, 2005). These experiments were able to capitalize on the well-defined circuitry to show that the interpositus nucleus was a locus of plasticity essential for acquiring and retaining a CS-US association capable of supporting conditioned blinks in response to the CS in a manner that was dissociable from the ability to emit unconditioned blinks in response to the US. Additional studies have indicated that the cerebellar cortex is an important part of the circuitry for modulating the timing of the conditioned response to ensure that it is adaptive.

Delineation of the functional circuitry of eyeblink classical conditioning, along with results from other paradigms and from computer simulations, has provided a view of the cerebellum and as a key part of a circuit capable of finely recalibrating the sub-second timing of outputs based on the precise timing of inputs (Ohyama, Nores, Murphy, & Mauk, 2003). By this view, the "error" signal of a puff of air to an open eye is used to recalibrate the timing of an eyeblink with the help of the predictive value of the CS so that the next puff might be received by a closed eye. The capacity of the cerebellum and associated circuitry to recalibrate sub-second timing may contribute in a similar manner to other examples of learning, such as in the recalibration of the timing of contraction of opposing arm muscles in executing coordinated reaching movements such that a missed object can be successfully grasped.

Summary of cerebellum

The circuitry essential for eyeblink classical conditioning has been identified and provides an experimentally tractable opportunity to investigate what may be generalized learning processes supported by the cerebellum. These processes may contribute generally to recalibrating the timing of outputs in sensorimotor learning or even more generally to mental processing involved in precise timing.

Amygdala

Anatomy

Two major areas of the amygdala involved in emotional memory are the basolateral complex of the amygdala (BLA; includes the lateral, basolateral, and basomedial nuclei) and the central nucleus of the amygdala (CEA; Fig. 48.14B). The BLA receives input from widespread cortical areas as well as from sensory nuclei of the thalamus and thus has access to higher-level information from association areas as well as lower-level sensory information. In addition, the BLA has reciprocal connections with other

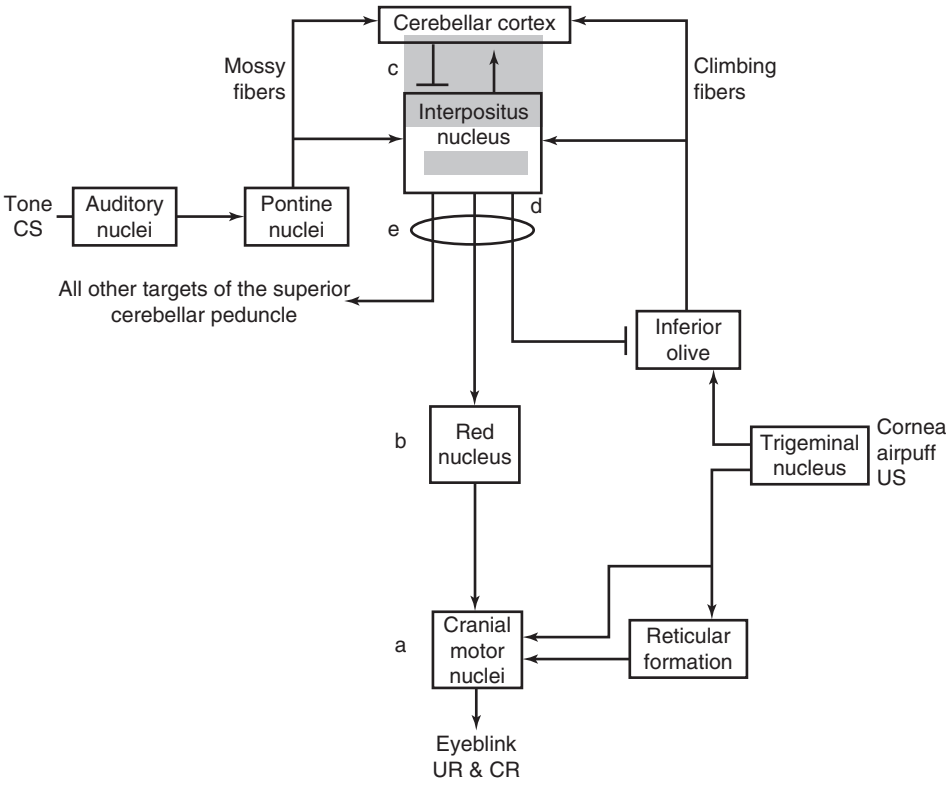


FIGURE 48.13 A schematic diagram of principal pathways involved in classical conditioning of the eyeblink reflex. The role of structures at points a–e has been studied using reversible inactivation with a local anesthetic. Inactivation at point "c" (shaded areas) prevents learning, whereas inactivation at "a," "b," "d," or "e" prevents the behavioral response during inactivation, but does not block learning. *From Thompson and Kim (1996).*

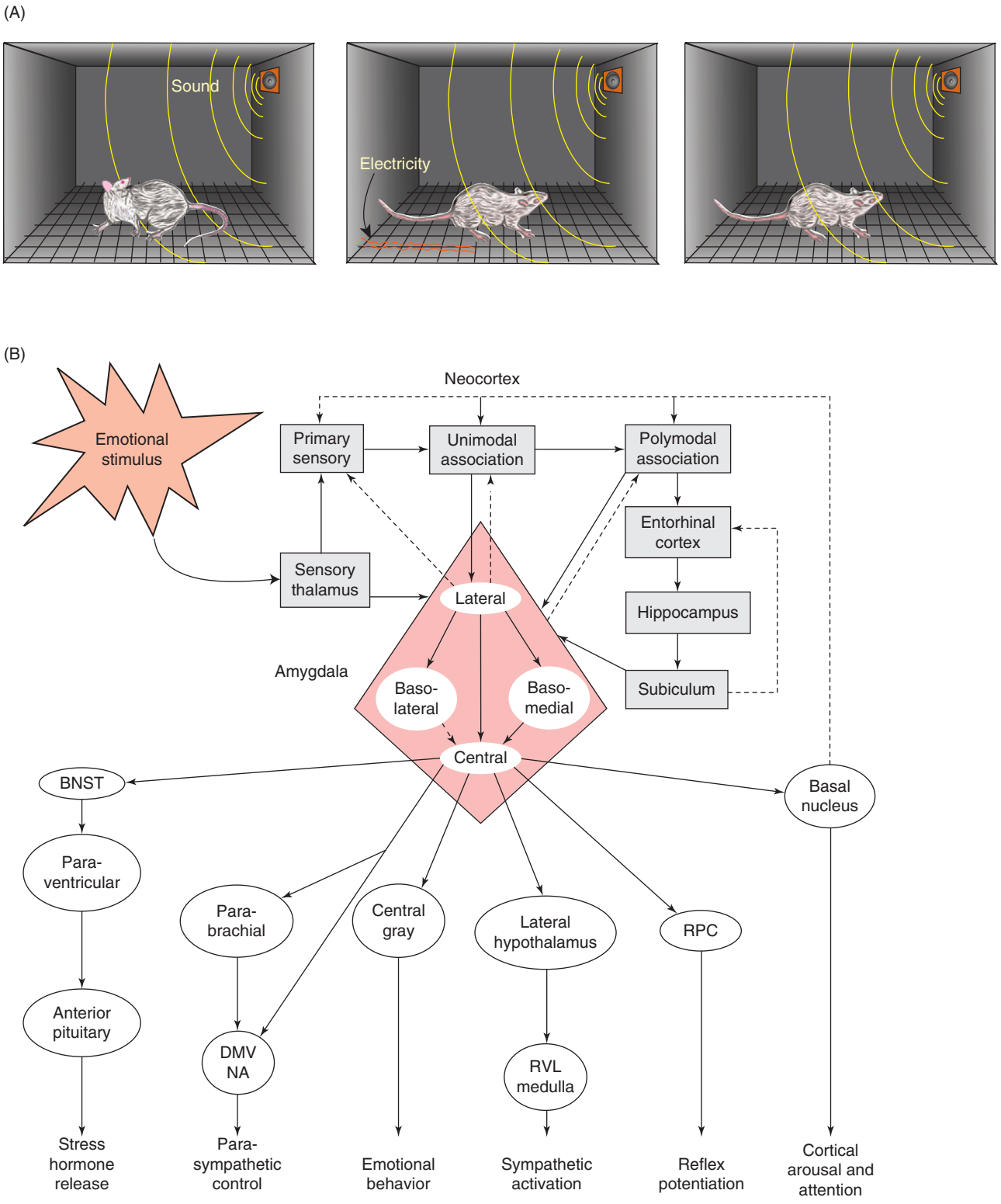


FIGURE 48.14 Fear conditioning. (A) Prior to training, the tone produces a transient orienting response. During training the tone is followed by a brief foot shock. Following training, the rat is reintroduced into the chamber and freezes when the tone is presented. (B) Anatomical pathways that mediate fear conditioning. A hierarchy of sensory inputs converges on the lateral amygdala nucleus, which projects to other amygdala nuclei and then to the central nucleus, which send outputs to several effector systems for emotional responses. BNST, bed nucleus of the stria terminalis; DMV, dorsal nucleus of the vagus; NA, nucleus ambiguus; RPC, nucleus reticularis pontis oralis; RVL, rostral ventral nucleus of the medulla. *From LeDoux (1995).*

brain systems, including the hippocampal and striatal memory systems. Areas within the BLA project to the central nucleus, which is the source of outputs to subcortical areas controlling a broad range of fear-related behaviors, including autonomic (e.g., changes in heart rate, blood pressure, sweating, or hormone release) and motor responses (e.g., fear-related freezing behavior).

The amygdala contributes to emotion in several ways, including mediating emotional influences on attention and perception and regulating emotional responses (Phelps & LeDoux, 2005). The topic of the present section is the two ways in which the amygdala contributes to emotional memory. First, the amygdala supports the acquisition of emotional dispositions towards stimuli. Second, the amygdala mediates the influence of emotion on the consolidation of memory in other memory systems.

Role of the Amygdala in Acquiring Emotional Dispositions to Stimuli

A particularly productive approach to studying the role of the amygdala in memory has been to focus on the circuitry that supports acquisition of fearful responses to simple auditory or visual stimuli (Fanselow & Gale, 2003; Lang & Davis, 2006; Paré, Quirk, & Ledoux, 2004). In one frequently used task (Fig. 48.14A), rats are initially habituated to a chamber and are then presented with multiple pairings of a tone (the CS) that terminates with a brief electric shock delivered through the floor of the chamber (the US), which typically results in an unconditioned jumping response (the UR). Subsequently, conditioned fear elicited by the tone is assessed by measuring autonomic responses or motor responses (the CR), such as freezing behavior. Rats with lesions in the BLA show markedly reduced conditioned fear responses to the tone but still show normal unconditioned fear responses to the shock itself. Rats with lesions to the central nucleus show reduced fear responses to both conditioned and unconditioned stimuli. Thus, the BLA appears to be important for acquiring an aversion to the tone, whereas the central nucleus appears to be additionally important for producing fearful responses in general.

A study in which fear conditioning was adapted for use in patients with damage to the amygdala or the hippocampus verified that the amygdala is crucial for acquiring conditioned fear responses and found further that these learned aversions are independent of declarative memory for the learning event (Bechera et al., 1995). In addition, functional imaging studies in healthy individuals have found increased activity in the amygdala for both positive and negative stimuli as compared to neutral stimuli (Vytal & Hamann, 2010), consistent with the view from experimental animals that, although its involvement in fear learning may be especially robust, the amygdala is important for acquiring both aversions and preferences to previously neutral stimuli (Morrison & Salzman, 2010).

Similar to the example of eyeblink classical conditioning and the cerebellum, identifying the confluence of CS and US pathways in the amygdala has provided opportunities for exploring the neurobiological details of emotional memory at a key locus of plasticity. These opportunities are also relevant to mental health issues including anxiety, phobias, and post-traumatic stress disorder, and a recent area of interest in emotional memory has been how the brain supports relearning that a stimulus no longer predicts a fear-inducing outcome (Jovanovic & Ressler, 2010). These studies suggest that the medial prefrontal cortex may play an important role by inhibiting activity in the amygdala in order to suppress previously learned fear responses and by mediating plasticity so that a new non-fearful disposition can be associated with the stimulus (Quirk & Mueller, 2008; Quirk et al., 2010).

Role of the Amygdala in the Modulation of Memory

Emotion can enhance memory, and evidence from a large number of studies in experimental animals has indicated that the amygdala, in particular the BLA, mediates this memory enhancing effect (McGaugh, 2004). In particular, emotionally arousing events result in the release of epinephrine and glucocorticoids by the adrenal glands, which in turn result in the release of norepinephrine in the amygdala (Fig. 48.15). The release of norepinephrine

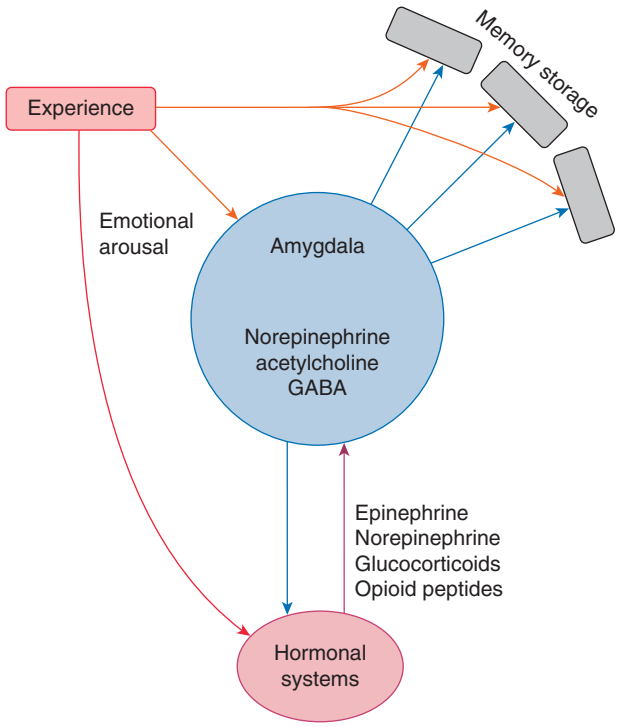


FIGURE 48.15 A schematic representation of how hormonal systems and the amygdala complex can modulate the storage of memory for emotionally arousing events through influences on other brain systems. See text for details. *From McGaugh, Introini-Collison, Cahill, Kim, and Liang (1992).*

leads to an increase in activity in the amygdala, which is thought to influence consolidation of memory in the other parts of the brain, either directly, through connections of the amygdala with the striatum, hippocampus, and cortex, or indirectly, through connections with nucleus basalis (which innervates much of the cortex). Data in support of this model include the findings that damage to the BLA or infusions of drugs in the BLA that interfere with norepinephrine block the memory enhancing effect of emotional arousal. In addition, infusion of norepinephrine directly into the BLA following a learning session can enhance subsequent memory, even in instances such as object recognition memory in which emotional arousal ordinarily does not make an essential contribution to the memory (Roozendaal, Castello, Vedana, Barsegyan, & McGaugh, 2008). Furthermore, the modulatory influence of the BLA seems to be limited to the time of the learning event as well as a short period of time after. Manipulations of the amygdala during subsequent testing of retention typically do not influence memory performance. Thus, the amygdala is important in modulating storage in emotionally stressful situations, but not in the maintenance or retrieval of the memory that has been modulated.

The amygdala has been strongly implicated in the enhancement of memory for emotional material in humans as well as in experimental animals. In one study, a patient with damage to the amygdala was tested for memory of an emotional story along with a group of healthy individuals (Adolphs, Cahill, Schul, & Babinsky, 1997). The participants watched a series of slides and listened to an accompanying narrative that told a story about a mother and son. One portion of the story involved a traumatic accident, and all participants rated this portion as being strongly emotional. Compared to the healthy individuals, the patient with damage to the amygdala failed to show enhancement of memory for the emotional part of the story. However, the patient performed as well as others on neutral story material. In brain imaging studies with healthy individuals, the amygdala was activated during the viewing of emotional material that elicited either positive or negative emotional reactions, and this activation was related to the likelihood that participants were able to subsequently remember the emotional material (Cahill et al., 1996; Canli, Zhao, Brewer, Gabrieli, & Cahill, 2000; Hamann, Ely, Grafton, & Kilts, 1999). In these studies, there was no correlation between activation of the amygdala and subsequent memory performance for neutral material.

Summary of Amygdala

The amygdala is well situated anatomically to attach emotional (positive and negative) dispositions to a broad range of stimuli and to mediate the acquisition of such dispositions in the absence of conscious recollection of the circumstances of the emotional experience. In addition, components of the amygdala also mediate the modulation of memory storage during and after emotional events. These two roles in memory could be brought together in situations in which, for example, a conditioned fear response to a stimulus leads to emotional arousal and to modulation of declarative memory for the event.

Cerebral Cortex

Anatomy

The posterior half of the cerebral cortex contains many functionally distinct areas that are organized into hierarchies of serial and parallel processing for each sensory modality (see Chapter 2). The anterior cerebral cortex contains a similar hierarchy of motor areas, as well as association areas in the prefrontal cortex involved with motor planning, higher order cognition, and working memory (Fuster, 2001; see Chapter 50). Additionally, information from different modalities converges on "association" areas in the parietal and temporal lobes.

Some of the examples provided so far, such as standard eyeblink classical conditioning, illustrate that some simple forms of learning can circumvent the cerebral cortex. Nevertheless, for all the memory systems discussed in this chapter, including the cerebellum, the cerebral cortex provides important access to incoming information as well as output routes to influence behavior. Yet plasticity in the cerebral cortex also directly supports types of memory that do not require components of the other memory systems. For each area of the cortex, the modification of its information processing circuitry through alterations in synaptic connectivity and membrane excitability underlies its direct participation in memory. These alterations can be dramatic, such as in examples of experience-dependent reorganization that occurs in perceptual or motor learning, but can also be more subtle, such as in the phenomenon of repetition priming.

Perceptual and Motor Learning

Extended experience with a particular set of stimuli or particular set of movements can result in the modification of the cortical areas important for perceiving those stimuli or executing those movements. In particular, studies in a variety of experimental animals have used single-neuron recordings to identify the stimulus-selectivity or movement-selectivity of individual neurons and to map their location in the cortex. A common finding in these studies is that training on a particular perceptual or motor task, such as a fine-grained auditory or tactile discrimination or a skilled reaching task, leads to an increase in the area of the cortex containing neurons that are selective for the stimuli or movements used in the task (Buonomano &

Merzenich, 1998; Monfils, Plautz, & Kleim, 2005; Weinberger, 2007). In some of these studies, cortical changes were correlated with task performance, suggesting that these cortical alterations were an important part of the acquired memory.

Functional imaging studies in humans have also suggested that the cortical areas important for perceiving the stimuli or executing the actions are also important areas of plasticity. For example, in one study participants were scanned while viewing pictures of nonsense objects both before and after performing a perceptual discrimination task that included half of the objects (Op de Beeck, Baker, DiCarlo, & Kanwisher, 2006). Following the perceptual discrimination task, the pattern of brain activity in object-responsive areas in visual cortex had changed for the objects included in the discrimination task but did not differ for the objects not included in the discrimination task. Thus, it appears that many of the areas in the cerebral cortex that are specialized for specific perceptual or information processing functions—from areas involved in perceiving simple tones to areas involved in perceiving complex objects—are capable of supporting perceptual learning for those stimuli.

Repetition Priming

Another example of memory thought to be supported directly by the cerebral cortex is the phenomenon known as repetition priming (or priming). Priming involves repeating previously presented items, either unaltered or as fragments of the original (e.g., word fragments or briefly flashed pictures), and measuring the increased fluency with which those items are identified as compared to items not repeated. For example, in one task, repeated and unrepeated words are flashed on a screen so briefly that many of the words are unreadable, and the measure of priming is the increased likelihood of identifying repeated words as compared with unrepeated words. Studies with amnesic patients and healthy individuals have found that successful priming is typically independent of the ability to recognize which items were repeated (Stark & Squire, 2000). Further, functional imaging studies in healthy individuals suggest that priming often results in reduced activity in cortical areas involved in initially processing the stimuli (Schacter, Wig, & Stevens, 2007), suggesting that less neural activity is required to identify words recently processed. These findings suggest that the neuronal basis of priming is an increase in the efficiency and bias in the direction of cortical sensory processing associated with perceptual identification.

Summary of Cerebral Cortex

The cerebral cortex provides a major route of information flow into and out of all of the memory systems discussed in this chapter. However, the cortex is more than just a way station. It is highly plastic, and distinct cortical areas are reorganized or biased by specific training experiences to meet demands for specific types of perceptual or motor representations.

BEHAVIOR SUPPORTED BY MULTIPLE MEMORY SYSTEMS

The discussion of memory systems so far has focused on dissociations between systems in illustrating the unique contributions of the central structure in each system. However, behavior is a product of the entire nervous system and is typically guided by contributions of more than one form of memory. For example, training in a sport normally results in declarative memory of instructions from one's coach but also leads to motor skill learning from repetitive practice. The memory for the instructions would depend on the hippocampus and various aspects of the skill would depend on the striatum, cerebellum, and motor cortex. The amygdala would likely also contribute during moments of emotional arousal. The following section considers emerging data regarding how the memory systems discussed so far might interact and illustrates some progress in understanding how the distribution of memory throughout the brain ordinarily contributes to our day to day activities.

One example of a direct collaboration between two memory systems is the influence of the amygdala on hippocampus-dependent memory for emotion-inducing material. The section earlier in the chapter on the amygdala discussed how emotional arousal leads to the amygdala engaging in modulation of declarative memory, such that stimuli that elicit either positive or negative emotional responses tend to be remembered better than neutral stimuli. This process is thought to result from amygdalar projections modulating synaptic plasticity either in the hippocampus and parahippocampal region or in cortical areas fundamental to the hippocampal memory system. It has been suggested that the amygdala also modulates other memory systems, such as the striatal memory system, and that the amygdala may play a generalized role in memory modulation for emotion inducing stimuli in addition to its role in attaching positive or negative dispositions to stimuli (McGaugh, 2004).

Another example comes from an indirect collaboration between two very dissimilar memory systems, the hippocampal memory system and the cerebellar memory system. The section earlier in the chapter on the cerebellum detailed that the essential circuitry for standard eyeblink classical conditioning involved no forebrain structures and instead centered on deep cerebellar nuclei. Yet the situation changes if the parameters for standard eyeblink classical conditioning, referred to as delay eyeblink conditioning, are altered slightly such that the tone that normally

overlaps with the airpuff is moved earlier in time such that it terminates as little as a half second prior to the onset of the airpuff, a version called trace conditioning (Christian & Thompson, 2003). Trace eyeblink conditioning depends on the same cerebellar and hindbrain circuitry as delay eyeblink conditioning, but trace conditioning depends additionally on forebrain structures including the hippocampus and prefrontal cortex. In studies with healthy human participants, trace (but not delay) eyeblink conditioning was found to relate closely to participants' knowledge about the relationship between the tone and the airpuff (Clark, Manns, & Squire, 2002). A possible explanation of these findings is that the silent interval that separates the tone and airpuff may surpass the learning abilities of the cerebellum and may cause it to require inputs from additional forebrain structures, such as the hippocampal system via the prefrontal cortex (Kalmbach, Ohyama, Kreider, Riusech, & Mauk, 2009), to bridge that brief interval. The growing understanding of the relationship between hippocampus and cerebellum illustrates the tendency to collaborate between even the most dissimilar memory systems, and exemplifies the progress being made towards understanding how memory systems interact in general.

In other examples, memory systems may not directly collaborate but may operate in parallel. For example, the previously discussed studies contrasting the role of the striatum and hippocampus in supporting "response" learning and "place" learning, respectively, illustrate how the memory supported by different brain systems may not converge. Yet in the example of practicing a sport, the hippocampus-dependent memory for a coach's instructions and the striatal "response" learning would likely support a common behavior. In these instances, it is thought that the striatum and hippocampus contribute to behavior in parallel rather than directly interacting.

CONCLUSION

Several brain systems support distinct forms of memory characterized by unique operating features and ways in which memory is expressed. The hippocampus supports rapidly-acquired memory that is experienced as conscious recollection and that can be expressed flexibly. The striatum supports acquisition of inclinations to make situational responses. The cerebellum supports learning related to adaptive timing and motor reflexes. The amygdala supports acquisition of emotional dispositions towards stimuli and also mediates the role of emotion arousal in modulating other forms of memory. The cerebral cortex participates in each of these types of memory, and mediates perceptual learning, motor learning, and the phenomenon of priming. These systems also often operate in concert to support the various ways that behavior and mental life can be modified by experience.

References

Adolphs, R., Cahill, L., Schul, R., & Babinsky, R. (1997). Impaired declarative memory for emotional material following bilateral amygdala damage in humans. *Learning and Memory, 4,* 291–300.

Alvarez, P., & Squire, L. R. (1994). Memory consolidation and the medial temporal lobe: A simple network model. *Proceeding of the National Academy Science, 15,* 7041–7045.

Amaral, D. G., & Witter, M. P. (2004). Hippocampal formation. In G. Paxinos (Ed.), *The rat nervous system*, (3rd ed., pp. 635–704). San Diego: Elsevier.

Bayley, P. J., & Squire, L. R. (2002). Medial temporal lobe amnesia: Gradual acquisition of factual information by nondeclarative memory. *Journal of Neuroscience, 13,* 5741–5748.

Bayley, P. J., Frascino, J. C., & Squire, L. R. (2005). Robust habit learning in the absence of awareness and independent of the medial temporal lobe. *Nature, 7050,* 550–553.

Bechera, A., Tranel, D., Hanna, D., Adolphs, R., Rockland, C., & Damasio, A. R. (1995). Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. *Science, 269,* 1115–1118.

Bi, G. Q., & Poo, M. M. (1998). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience, 18,* 10464–10472.

Boyden, E. S., Katoh, A., & Raymond, J. L. (2004). Cerebellum-dependent learning: The role of multiple plasticity mechanisms. *Annual Review of Neuroscience, 27,* 581–609.

Buffalo, E. A., Bellgowan, P. S., & Martin, A. (2006). Distinct roles for medial temporal lobe structures in memory for objects and their locations. *Learning and Memory, 13,* 638–643.

Bunsey, M., & Eichenbaum, H. (1996). Conservation of hippocampal memory function in rats and humans. *Nature, 379,* 255–257.

Buonomano, D. V., & Merzenich, M. M. (1998). Cortical plasticity: From synapses to maps. *Annual Review of Neuroscience, 21,* 149–186.

Burwell, R. D., Witter, M. P., & Amaral, D. G. (1995). Perirhinal and postrhinal cortices in the rat: A review of the neuroanatomical literature and comparison with findings from the monkey brain. *Hippocampus, 5,* 390–408.

Cahill, L., Haier, R. J., Fallon, J., Alkire, M. T., Tang, C., Keator, D., …, McGaugh, J. L. (1996). Amygdala activity at encoding correlated with long-term, free recall of emotional information. *Proceeding of the National Academy of Science USA, 93,* 8016–8021.

Canli, T., Zhao, Z., Brewer, J., Gabrieli, J. D., & Cahill, L. (2000). Event-related activation in the human amygdala associates with later memory for individual emotional experience. *Journal of Neuroscience, 20*(RC99), 1–5.

Christian, K. M., & Thompson, R. F. (2003). Neural substrates of eyeblink conditioning: Acquisition and retention. *Learning and Memory, 10,* 427–455.

Clark, R. E., Manns, J. R., & Squire, L. R. (2002). Classical conditioning, awareness, and brain systems. *Trends in Cognitive Sciences, 6,* 524–531.

Colgin, L. L., Denninger, T., Fyhn, M., Hafting, T., Bonnevie, T., Jensen, O., …, Moser, E. I. (2009). Frequency of gamma oscillations routes flow of information in the hippocampus. *Nature, 462,* 353–357.

Derdikman, D., & Moser, E. I. (2010). A manifold of spatial maps in the brain. *Trends in Cognitive Sciences, 14,* 561–569.

Diba, K., & Buzsáki, G. (2007). Forward and reverse hippocampal place-cell sequences during ripples. *Nature Neuroscience, 10,* 1241–1242.

Eichenbaum, H. (2001). The hippocampus and declarative memory: Cognitive mechanisms and neural codes. *Behavioural Brain Research, 127,* 199–207.

Fanselow, M. S., & Gale, G. D. (2003). The amygdala, fear, and memory. *Annals of the New York Academy Science, 985,* 125–134.

Fell, J., Klaver, P., Lehnertz, K., Grunwald, T., Schaller, C., Elger, C. E., & Fernandez, G. (2001). Human memory formation is accompanied by rhinal-hippocampal coupling and decoupling. *Nature Neuroscience, 4,* 1259–1264.

Foster, D. J., & Wilson, M. A. (2006). Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature, 440,* 680–683.

Fuentemilla, L., Penny, W. D., Cashdollar, N., Bunzeck, N., & Duzel, E. (2010). Theta-coupled periodic replay in working memory. *Current Biology, 20,* 606–612.

Fuster, J. M. (2001). The prefrontal cortex—An update: time is of the essence. *Neuron, 30,* 319–333.

Gelbard-Sagiv, H., Mukamel, R., Harel, M., Malach, R., & Fried, I. (2008). Internally generated reactivation of single neurons in human hippocampus during free recall. *Science, 322,* 96–101.

Graybiel, A. M. (2008). Habits, rituals, and the evaluative brain. *Annual Review of Neuroscience, 31,* 359–387.

Hafting, T., Fyhn, M., Molden, S., Moser, M. B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature, 436,* 801–806.

Hamann, S. B., Ely, T. D., Grafton, S. T., & Kilts, C. D. (1999). Amygdala activity related to enhanced memory for pleasant and aversive stimuli. *Nature Neuroscience, 2,* 289–293.

Hasselmo, M. E., Bodelon, C, & Wyble, B. C. (2002). A proposed function for hippocampal theta rhythm: Separate phases of encoding and retrieval enhance reversal of prior learning. *Neural Computation, 14,* 793–817.

Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.

Jovanovic, T., & Ressler, K. J. (2010). How the neurocircuitry and genetics of fear inhibition may inform our understanding of PTSD. *American Journal of Psychiatry, 167,* 648–662.

Jutras, M. J., Fries, P., & Buffalo, E. A. (2009). Gamma-band synchronization in the macaque hippocampus and memory formation. *Journal of Neuroscience, 29,* 12521–12531.

Jutras, M. J., & Buffalo, E. A. (2010). Recognition memory signals in the macaque hippocampus. *PNAS, 107,* 401–406.

Kalmbach, B. E., Ohyama, T., Kreider, J. C., Riusech, F., & Mauk, M. D. (2009). Interactions between prefrontal cortex and cerebellum revealed by trace eyelid conditioning. *Learning and Memory, 16,* 86–95.

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science, 273,* 1399–1401.

Kreiman, G., Koch, C., & Fried, I. (2000). Category-specific visual responses of single neurons in the human medial temporal lobe. *Nature Neuroscience, 3,* 946–953.

Lang, P. J., & Davis, M. (2006). Emotion, motivation, and the brain: reflex foundations in animal and human research. *Progress in Brain Research, 156,* 3–29.

LeDoux, J. E. (1995). Emotion: Clues from the brain. *Annual Review of Psychology, 46,* 209–235.

Litman, L., Awipi, T., & Davachi, L. (2009). Category-specificity in the human medial temporal lobe cortex. *Hippocampus, 19,* 308–319.

Manns, J. R., & Eichenbaum, H. (2007). Evolution of the hippocampus. In L. Krubitzer & J. Kaas (Eds.), *Evolution of nervous systems: The evolution of nervous systems in mammals* (Vol. 4). London: Elsevier.

Manns, J. R., Zilli, E. A., Ong, K. C., Hasselmo, M. E., & Eichenbaum, H. (2007). Hippocampal CA1 spiking during encoding and retrieval: Relation to theta phase. *Learning and Memory, 87,* 9–20.

Marshall, L., Helgadottir, H., Molle, M., & Born, J. (2006). Boosting slow oscillations during sleep potentiates memory. *Nature, 444,* 610–613.

Mauk, M. D., & Thompson, R. F. (1987). Retention of classically conditioned eyelid responses following acute decerebration. *Brain Research, 403,* 89–95.

Maviel, T., Durkin, T. P., Menzaghi, F., & Bontempi, B. (2004). Sites of neocortical reorganization critical for remote spatial memory. *Science, 305,* 96–99.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102,* 419–457.

McGaugh, J. L. (2004). The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annual Review of Neuroscience, 27,* 1–28.

McGaugh, J. L., Introini-Collison, I. B., Cahill, L., Kim, M., & Liang, K. C. (1992). Involvement of the amygdala in neuromodulatory influences on memory storage. In J. P. Aggleton (Ed.), *The amgydala: Neurobiological aspects of emotion, memory, and mental dysfunction* (pp. 431–451). New York: Wiley-Liss.

Middleton, F. A., & Strick, P. L. (1998). Cerebellar output: Motor and cognitive channels. *Trend in Cognitive Science, 2,* 348–354.

Milner, B. (1962). Les troubles de la memoire accompagnant les lesions hippocampiques bilaterales. In P. Passouant (Ed.), *Physiologie Del'-hippocampe, Colloques Internationaux no. 107* (pp. 257). Paris: CNRS. [English translation 1965. In B. Milner & S. Glickman (eds.), *Cognitive processes and the brain* (p. 97). Princeton, NJ: Van Nostrand.]

Mishkin, M. (1978). Memory in monkeys severely impaired by combined but not separate removal of the amygdala and hippocampus. *Nature, 273,* 297–298.

Monfils, M. H., Plautz, E. J., & Kleim, J. A. (2005). In search of the motor engram: motor map plasticity as a mechanism for encoding motor experience. *Neuroscientist, 11,* 471–483.

Morris, R. G. M., Garrud, P., Rawlins, J. N. P., & O'Keefe, J. (1982). Place navigation impaired in rats with hippocampal lesions. *Nature, 297,* 681–683.

Morrison, S. E., & Salzman, C. D. (2010). Re-valuing the amygdala. *Current Opinion Neurobiology, 20,* 221–230.

Nakashiba, T., Young, J. Z., McHugh, T. J., Buhl, D. L., & Tonegawa, S. (2008). Transgenic inhibition of synaptic transmission reveals role of CA3 output in hippocampal learning. *Science, 319,* 1260–1264.

Naya, Y., & Suzuki, W. A. (2011). Integrating what and when across the primate medial temporal lobe. *Science, 333,* 773–776.

Ohyama, T., Nores, W. L., Murphy, M., & Mauk, M. D. (2003). What the cerebellum computes. *Trends in Neuroscience, 26,* 222–227.

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research, 34,* 171–175.

Op de Beeck, H. P., Baker, C. I., DiCarlo, J. J., & Kanwisher, N. G. (2006). Discrimination training alters object representations in human extrastriate cortex. *Journal of Neuroscience, 26,* 13025–13036.

Packard, M. G., & McGaugh, J. L. (1996). Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. *Neurobiology of Learning and Memory, 65,* 65–72.

Paré, D., Quirk, G. J., & Ledoux, J. E. (2004). New vistas on amygdala networks in conditioned fear. *Journal of Neurophysiology, 92,* 1–9.

Phelps, E. A., & LeDoux, J. E. (2005). Contributions of the amygdala to emotion processing: From animal models to human behavior. *Neuron, 48,* 175–187.

Pihlajamaki, M., Tanila, H., Kononen, M., Hanninen, T., Hamalainen, A., Soininen, H., & Aronen, H. J. (2004). Visual presentation of novel objects and new spatial arrangements of objects differentially activates the medial temporal lobe subareas in humans. *European Journal of Neuroscience, 19,* 1939–1949.

Polyn, S. M., Natu, V. S., Cohen, J. D., & Norman, K. A. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science, 310,* 1963–1966.

Quirk, G. J., & Mueller, D. (2008). Neural mechanisms of extinction learning and retrieval. *Neuropsychopharmacology, 33,* 56–72.

Quirk, G. J., Paré, D., Richardson, R., Herry, C., Monfils, M. H., Schiller, D., & Vicentic, A. (2010). Erasing fear memories with extinction training. *Journal of Neuroscience, 30,* 14993–14997.

Rolls, E. T. (2010). A computational theory of episodic memory formation in the hippocampus. *Behavioual Brain Research, 215,* 180–196.

Roozendaal, B., Castello, N. A., Vedana, G., Barsegyan, A., & McGaugh, J. L. (2008). Noradrenergic activation of the basolateral amygdala modulates consolidation of object recognition memory. *Neurobiology of Learning and Memory, 90,* 576–579.

Rutishauser, U., Mamelak, A. N., & Schuman, E. M. (2006). Single-trial learning of novel stimuli by individual neurons of the human hippocampus-amygdala complex. *Neuron, 49,* 805–813.

Rutishauser, U., Ross, I. B., Mamelak, A. N., & Schuman, E. M. (2010). Human memory strength is predicted by theta-frequency phase-locking of single neurons. *Nature, 464,* 903–907.

Schacter, D. L., Wig, G. S., & Stevens, W. D. (2007). Reductions in cortical activity during priming. *Current Opinion in Neurobiology, 17,* 171–176.

Schultz, W. (2007). Behavioral dopamine signals. *Trends in Neuroscience, 30,* 203–210.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, Psychiatry, 20,* 11–21.

Sederberg, P. B., Schulze-Bonhage, A., Madsen, J. R., Bromfield, E. B., McCarthy, D. C., Brandt, A., …, Kahana, M. J. (2007). Hipocampal and neocortical gamma oscillations predict memory formation in humans. *Cerebral Cortex, 17,* 1190–1196.

Squire, L. R. (2009). The legacy of patient H.M. for neuroscience. *Neuron, 61,* 6–9.

Squire, L. R., Clark, R. E., & Knowlton, B. J. (2001). Retrograde amnesia. *Hippocampus, 11,* 50–55.

Stark, C. E., & Squire, L. R. (2000). Recognition memory and familiarity judgments in severe amnesia: no evidence for a contribution of repetition priming. *Behavioral Neuroscience, 114,* 459–467.

Stefanacci, L., Buffalo, E. A., Schmolck, H., & Squire, L. R. (2000). Profound amnesia after damage to the medial temporal lobe: A neuroanatomical and neuropsychological profile of patient E. P. *Journal of Neuroscience, 20,* 7024–7036.

Stern, C. E., Corkin, S., Gonzalez, R. G., Guimaraes, A. R., Baker, J. R., Jennings, P. J., …, Rosen, B. R. (1996). The hippocampal formation participates in novel picture encoding: Evidence from functional magnetic resonance imaging. *Proceeding of the National Academy Science USA, 93,* 8660–8665.

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature, 437,* 1272–1278.

Sutherland, G. R., & McNaughton, B. (2000). Memory trace reactivation in hippocampal and neocortical neuronal ensembles. *Current Opinion in Neurobiology, 10,* 180–186.

Suzuki, W. A., & Amaral, D. G. (1994). Perirhinal and parahippocampal cortices of the macaque monkey: cortical afferents. *Journal of Comparative Neurology, 350,* 497–533.

Suzuki, W. A. (1996). Neuroanatomy of the monkey entorhinal, perirhinal, and parahioppocampal cortices: Organization of cortical inputs and interconnections with amygdala and striatum. *Seminars in Neuroscience, 8,* 3–12.

Teng, E., & Squire, L. R. (1999). Memory for places learned long ago is intact after hippocampal damage. *Nature, 400,* 675–677.

Thompson, R. F. (2005). In search of memory traces. *Annual Review of Psychology, 56,* 1–23.

Thompson, R. F., & Kim, J. J. (1996). Memory systems in the brain and localization of a memory. *Proceedings of the National Academy of Science USA, 93,* 13438–13444.

van der Meer, M. A., Johnson, A., Schmitzer-Torbert, N. C., & Redish, A. D. (2010). Triple dissociation of information processing in dorsal striatum, ventral striatum, and hippocampus on a learned spatial decision task. *Neuron, 67,* 25–32.

Vytal, K., & Hamann, S. (2010). Neuroimaging support for discrete neural correlates of basic emotions: a voxel-based meta-analysis. *Journal of Cognitive Neuroscience, 22,* 2864–2885.

Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M., …, Buckner, R. L. (1998). Building memories: Remembering and forgetting of verbal experiences as predicted by brain activity. *Science, 281,* 1188–1191.

Weinberger, N. M. (2007). Associative representational plasticity in the auditory cortex: A synthesis of two disciplines. *Learning and Memory, 14,* 1–16.

Wirth, S., Yanike, M., Frank, L. M., Smith, A. C., Brown, E. N., & Suzuki, W. A. (2003). Single neurons in the monkey hippocampus and learning of new associations. *Science, 300,* 1578–1581.

Witter, M. P., Naber, P. A., va Haeften, T., Machielsen, W. C., Rombouts, S. A., Barkhof, F., …, Lopes da Silva, F. H. (2000). Cortico-hippocampal communication by way of parallel parahippocampal-subicular pathways. *Hippocampus, 10,* 398–410.

Wood, E. R., Dudchenko, P. A., & Eichenbaum, H. (1999). The global record of memory in hippocampal neuronal activity. *Nature, 397,* 613–616.

Yassa, M. A., & Stark, C. E. L. (2011, July 23). Pattern separation in the hippocampus. *Trends in Neuroscience, 34,* 515–525.

Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Review Neuroscience, 7,* 464–476.

Zola-Morgan, S., & Squire, L. R. (1985). Medial temporal lesions in monkeys impair memory on a variety of tasks sensitive to human amnesia. *Behavioral Neuroscience, 99,* 22–34.

Zola-Morgan, S., Squire, L. R., & Amaral, D. G. (1986). Human amnesia and the medial temporal region: Enduring memory impairment following a bilateral lesion limited to field CA1 of the hippocampus. *Journal of Neuroscience, 6,* 2950–2967.

Zola, S. M., Squire, L. R., Teng, E., Stefanacci, L., Buffalo, E. A., & Clark, R. E. (2000). Impaired recognition memory in monkeys after damage limited to the hippocampal region. *Journal of Neuroscience, 20,* 451–463.

Suggested Readings

Anderson, P., Morris, R., Amaral, D., Bliss, T., & O'Keefe, J. (2007). *The hippocampus book*. New York: Oxford Univ. Press.

Buzsaki, G. (2006). *Rhythms of the brain*. New York: Oxford Univ. Press.

Dudai, Y. (2002). *Memory from a to z: Keywords, concepts, and beyond*. New York: Oxford Univ. Press.

Eichenbaum, H. (2008). *Learning & memory*. New York: W.W. Norton.

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. London: Oxford Univ. Press (Clarendon).

Schacter, D. L., & Tulving, E. (Eds.), (1994). *Memory systems*. Cambridge, MA: MIT Press.

Squire, L. R., & Kandel, E. R. (2008). *Memory: From mind to molecules* (2nd ed). Greenwood Village: Roberts & Company Publishers.

Squire, L. R. (2004). Memory systems of the brain: a brief history and current perspective. *Neurobiology of Learning and Memory, 82,* 171–177.

Steiner, H., & Tseng, K. Y. (Eds.), (2010). *Handbook of basal ganglia structure and function*. London: Academic Press.

Whalen, P. J., & Phelps, E. A. (Eds.), (2009). *The human amygdala*. New York: Guilford Press.

Joseph R. Manns and Elizabeth A. Buffalo

This page intentionally left blank

CHAPTER

49

Language

This chapter discusses the relationship of language to the brain. It begins with a discussion of animal communication, from which human language evolved and that provides a guide to the basic neural elements and processes that underlie our own system of communication. The next section of the chapter provides a brief summary of human language and its relation to the brain, relying on studies of patients with various types of neurological diseases and on brain imaging in normal subjects.

ANIMAL COMMUNICATION

Communication Is Important for Individual Survival

The ultimate goal of most animal communication, like that of behavior in general, is reproduction. Thus, signals that indicate the sender's species, gender, and degree of reproductive readiness account for the vast majority of natural communication. These are the messages being broadcast by chirping crickets, flashing fireflies, pheromone-releasing moths, and singing birds. Frequently, the mate-attraction call is also used to warn off potential rivals; for example, bird and cricket songs are also used to identify the boundaries of territories or personal space. Differences in signal quality are the usual basis of mate choice in species for which some degree of discrimination is evident. This competition for the attention of members of the opposite sex has led to the development of more conspicuous signals and, in many cases, to the evolution of displays, signaling morphology (i.e., sounds vs. shapes vs. color vs. movement vs. odors), and messages that go far beyond the basic needs of species and sexual identification (Gould & Gould, 1996).

Most animals are solitary except when mating; they abandon their eggs or larvae before the offspring are born. However, a number of species engage in some degree of parental care. For them, signals between parent and offspring are often very important. Most birds, for instance, have about two dozen calls that communicate mundane messages such as the need to eat, defecate, take cover, and so on. In cases in which both parents tend the young—the usual circumstance in birds, for instance—additional signals are required to agree on a nest site, synchronize brooding shifts, and guard the nestlings. Most primates also utilize two to three dozen signals for similar purposes.

The minority of species that are highly social have the most elaborate communication systems of all. They need messages for a variety of elements of social coordination, including, in many cases, group hunting or foraging, defense, and working out of a social hierarchy.

Animal Communication Strategies and Mechanisms

To ensure species specificity in mating, most species rely on more than one cue to identify a sexual partner. (Exceptions include some of the species that rely on pheromones.) Thus, multiple signals are sent, and a choice must be made between sending simultaneous messages and sending sequential messages (or a combination of the two). The sequential strategy has the advantage that the individual signals must be correct *and* the order must be appropriate. A female stickleback, for instance, requires the male to have a red ventral stripe, perform a zigzag dance, poke his nose in a nest, and then vibrate her abdomen; the odds of this concatenation of signals occurring together in this order by chance are remote. However, sequential signaling is time-consuming. A faster strategy is to provide all the cues in parallel, an approach that accepts the larger chance that these cues can occur together by chance. For example,

© 2013 Elsevier Inc. All rights reserved.

when a parent herring gull waves its bill in front of chicks to see if they need to be fed, the young simultaneously see a vertical beak, a red spot, and a horizontal motion, each of which is a discrete cue that combine in the mind of the chick to elicit pecking.

Innate and Learned Aspects of Animal Communication

Nearly all animal communication is innate: the sender produces the appropriate signal in the correct context even without any opportunity for learning, and it can be recognized for what it is by equally naive conspecifics. The basis of innate recognition appears to lie with feature detectors in the nervous systems—the inborn circuits that automatically isolate iconic visual or acoustic elements (Gould, 1982). The availability of many visual, auditory, tactile, and olfactory feature detectors, which can, in theory, be used in any specific combination or order, accounts for most of the diversity of animal communication, and reliance on these individual feature detectors, rather than pattern detection, accounts for its limitations. More complex behavior in relatively short-lived species, such as nest building in birds, is also usually innate, although some improvement with experience is also evident (Gould & Gould, 1999, 2007). Therefore, looking at how complex innate communication can be is useful in considering the extent to which as complex a system as human language must be learned.

In terms of its ability to communicate information, the most complex system of nonhuman communication known at present is the dance of honeybees. The system has some properties that are reminiscent of human language: it uses arbitrary conventions to describe objects distant in both space and time; that is, it does not reflect a real-time emotive readout, as might be the case when a primate gives an alarm call or grunts at a banana.

In terms of information content, a honeybee dance is second (albeit a very distant second) only to human speech, and so far as is known is far richer than the language of any primate. However, the dance language suffers from at least two limitations: it is a closed system in that there seems to be no way to introduce new conventions to deal with novel needs, and it is also graded; instead of discrete signals for different directions or distances, single components are varied over a range of values (angles and durations). The less complex but more flexible systems of birds and primates are more likely to illustrate the evolutionary precursors of speech.

Some birds have innate songs. In these species, individuals raised in isolation produce songs that are indistinguishable from their socially reared peers and respond to calls appropriately without prior experience. Chickens, doves, gulls, and ducks are familiar examples. Most songbirds, however, illustrate a different pattern: one that shows how potent a combination instinct and innately directed learning can be. Isolated chicks sing a schematic form of the species song, but the richness of a normal song is absent. Adult conspecifics can recognize innate songs as coming from members of their species, but in general these impoverished vocalizations produce lower levels of response. Typically, there is a sensitive period during which exposure to song must occur if it is going to have any impact (Fig. 49.1). In most species, there is a gap between this sensitive period and the process of overt song development—that is, practice and perfection of the adult song are based on the bird's memory of what it heard during its sensitive period (Gould & Marler, 1987).

Given a range of possible song models during isolated rearing, a chick selects an example from its own species and memorizes it. If it hears only songs of other species, the mature song is the unmodified innate song (Fig. 49.1). Thus there is an innate bias in the initial learning. Where this bias has been studied, it appears to depend on acoustic sign stimuli (i.e., species-specific "syllables"). Chicks are able to extract syllables of their species embedded in foreign songs, or scored in a way never found in their species (e.g., a syllable that is repeated at an accelerating rate presented to a species that sings syllables at a constant rate, or vice versa will be extracted and used in the species-typical manner).

Practice is essential in the normal development of birdsong, and part of this practice occurs in a babbling phase known as subsong, which begins at a species-typical age. A bird deafened after its sensitive period, but before it begins producing notes in preparation for singing (subsong), is unable to produce even an innate song (Fig. 49.1). During the earliest parts of subsong, birds try out a number of notes. These notes are typical of the species, but most are absent from the song they eventually sing. The learning process may involve producing each member of an innate repertoire of notes, listening to them, checking to see if they match any element in the memorized song, discarding the unnecessary ones, and rearranging, scoring, and modifying the others to produce a reasonable copy of the original song heard during the sensitive period.

There is some flexibility in song development. For example, when the chick has heard two very different specimens of its own species' song, it will often incorporate elements from each. Similarly, when the chick has been exposed to the sight of a singing conspecific and simultaneously the sound of a heterospecific song, it may pick out elements of the abnormal song and adapt them as best it can into its own species-specific organization.

Birdsong, therefore, depends on two processes that involve an interaction between innate capacities and learning: imprinting the song in memory and then learning to perform it. This system is flexible, but only within clear limits.

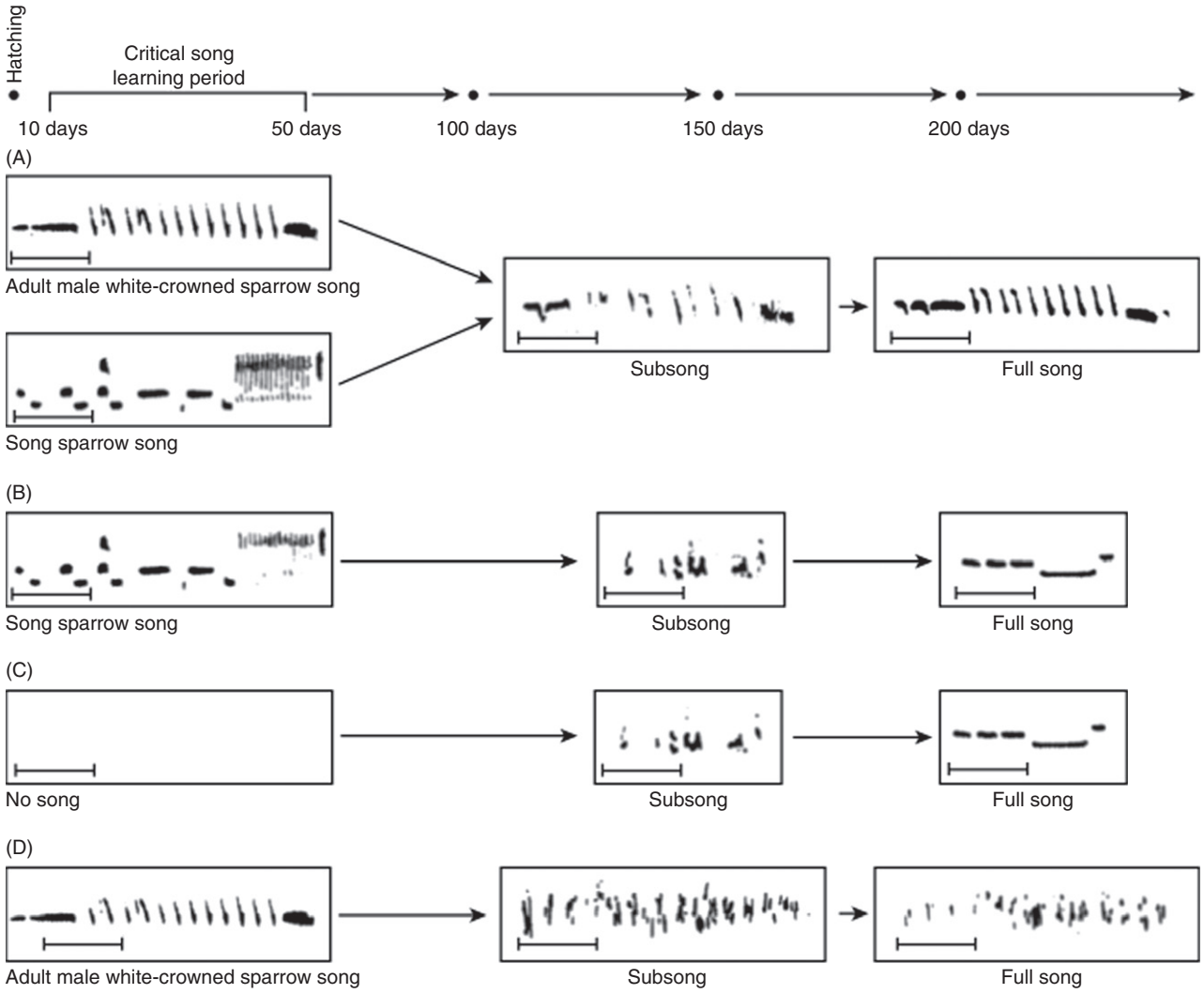


FIGURE 49.1 Birdsong development in most species is characterized by a sensitive period during which a song of the species must be heard. Later, during subsong, the bird practices making notes and assembles them into the correct order and pattern (A). Birds not allowed to hear their species' song sing a schematic version of the song (B and C); birds deafened before subsong cannot sing (D).

Communication Systems in Primates

Primates are the species with the closest evolutionary links to humans, and therefore the communication systems found in these animals are important to study for clues regarding the neural basis of human language.

Vervets, a species of monkey, provide a good case study. Vervets, like all social primates, have a large repertoire of innate calls used for social communication. Among these approximately three dozen signals are four alarm calls (Cheney & Seyfarth, 1990). In some parts of their range, one of the calls is specific for martial eagles; in another, the same call is used for certain hawks. In either case, the call causes monkeys to look up; those at the tops of trees drop to the interior, whereas those on the ground move into bushes or under trees. A second call, specific to leopards in one region and to other solitary hunters elsewhere, sends the warned individuals up to the tree tops. A third, specific for snakes, induces the other members of the troop to stand up and look around in the grass. A fourth call is heard in the presence of humans or group-hunting predators.

The development of calling is revealing. Young vervets appear to understand the class of animals each call refers to, but not the particular species that are dangerous. Thus, infants will give the eagle call to harmless vultures, storks, and even falling leaves, but not to snakes or leopards. Consequently, adults generally respond to the alarm calls of infants with a casual look around, followed by their own alarm if there is a genuine danger. Juveniles make fewer mistakes, and adult errors are confined to calls produced when the potential threat is so far away that human

observers require binoculars to identify the species. In short, young vervets seem to learn the details of how to apply an innate categorical vocabulary.

Nonhuman primates have virtually no voluntary control over respiration, the vocal tract, or vocalization, so this modality is not available for producing a large number of differentiated signals. Other modalities of communication, such as gesture, however, are also used for relatively restricted purposes, mostly related to items of immediate biological importance and social hierarchy. Recent detailed studies of gesture use in captive chimpanzees indicate that the meanings of particular gestures differ from one group to another.

Reseachers have attempted to teach chimpanzees to use a variety of communicative systems that bypass their vocal limitations, such as sign language and manipulation of object tokens, in an effort to determine to what extent these naturally occurring communication systems can be augmented. If anything, this work emphasizes the ways in which nonhuman primates are not capable of full human language. There are dramatic differences between trained chimpanzees and humans in terms of vocabulary size (chimps never get beyond what a 2-year-old can do), conceptual capacities, complex grammar, and so on, despite the fact that chimps have remarkable skills at social cognition.

Neural Structures for Animal Communication Systems

The neural mechanisms that underlie animal communication may give clues as to those that allow for human language. These systems of course differ in different phyla and species. This section briefly reviews aspects of the neural basis for birdsong and primate communication. It would be of great interest to report on the neural basis for the dance of the bees, but nothing is known about this subject.

Birdsong

Birdsong appears to be produced by a small number of nuclei (collections of nerve cells). In the canary, four important nuclei are involved: HVc, RA, DM, and nXIIts. Lesions in any of these nuclei destroy or seriously impair song production. These nuclei also respond when song is presented, and some neurons in some nuclei in some species (e.g., neurons in nXIIts in the zebra finch and the white-crowned sparrow) respond specifically to particular aspects of song. This indicates an overlap in the neural mechanisms involved in song perception and production, a feature that may carry over to humans as well.

Birdsong is produced by a structure between the trachea and the bronchi called the syrinx. There is one syrinx on each side of the trachea. In some species, song is entirely, or mainly, produced by one of these two structures; in other species, both syrinxes participate in song, with each producing different parts of the song. Each syrinx is controlled by structures in its corresponding side of the brain and, when one syrinx is responsible for song, these structures are larger on that side. The perceptual system is also duplicated on both sides of the brain, and, in some species, the two sides of the brain respond to different aspects of song.

In species in which only male birds sing (most species), song nuclei are much larger in males than in females ("sexual dimorphism"). This is not true in those species in which males and females perform duets. In general, there is a correlation between song nucleus size and song complexity in individual birds. Furthermore, song nuclei grow by a factor of almost 100% during the spring mating season (when birds sing) compared to the fall and winter (when they do not).

Changes in the size of song nuclei are related to testosterone production (they can be induced in females by testosterone injection in some circumstances) and by other factors usually related to day length—part of the birds' innate calendar. They are due to many changes in nerve cells: increase in the number of dendrites, increase in cell size, and, most interestingly, increase in the number of cells. Studies using labeled thymidine, which is taken up by dividing DNA, have shown neurogenesis in neurons in the HVc nucleus in relationship to periods of increased song, indicating that new nerve cells are formed. These neurons project to the RA nucleus, strongly suggesting that they play a role in song.

Primates

The neural basis for primate vocalizations is beginning to be charted. Unlike humans, vocalizations in primates do not appear to begin in a lateral cortical region but rather in the medial portion of the brain (the cingulate), a region involved in connecting basic instinctual to more advanced cognitive functions. Actual motor planning appears to involve mainly brainstem nuclei. On the perception side, there are neurons in the auditory cortex of some species of macaque monkeys that selectively respond to conspecific calls. These regions show some degree of functional asymmetry in some species, with responses from cells in the left hemisphere and not the right.

Possible Lessons from Nonhuman Species Regarding Language and Its Neural Basis

It is dangerous to infer too much about human language, or about how the human brain may support language, from studies of species that are very distant, such as birds. However, these species provide evidence for the possibility of certain aspects of language and its neural basis that could have developed several times in evolution or that could date to a distant common ancestor.

At a behavioral level, human speech and language could have developed from the repertoire of two to three dozen innate calls that are typical of birds and primates. In that case, we would expect to find that language contains elements that are innately recognized and that may be distinguished by the sorts of acoustic features that provide the basis for the innate discrimination of sign stimuli in other species. In fact, babies have an innate ability to discriminate a small number of sounds found in human languages, which function as acoustic sign stimuli. The evolution of human speech could also have retained the sorts of sensitive periods and innate learning biases so evident in song birds and vervet monkeys. In this case, we might expect to see a species-typical babbling phase, and perhaps a sensitive period for easy learning of new languages. If the evolution of speech and language preserves the use of preexisting (innate) forms, we might expect there to be innate aspects of human language, which would presumably be found in all languages and which might surface in situations where we see a "default grammar," not unlike the unlearned and impoverished songs of birds reared in isolation. Creoles potentially exemplify such cases. Creoles are languages developed by the children of people speaking different languages who are drawn together and communicate with very simple pidgin. They have a number of features not present in the languages spoken by the parents of the children who spoke the pidgin, which have been taken to be related to innately specified features of language (Bickerton, 1990).

From the neurological perspective, several features of the neural organization for communication systems appear to be found in humans in a way that is relevant to language. These features are the fact that there are specialized neural nuclei for song production; that the neural basis for song production is often asymmetric, with one hemisphere producing most of song or the two hemispheres producing different aspects of song; that production and perception of song use the same structures to some degree; and that there is a relationship of size to function. However, other important features of the neurobiology of birdsong, such as the neurogenesis that is important in the seasonal changes in song production, are not known to play a role in human speech and language. As for primates, the neural system responsible for call production is quite different from that in humans, being based in the cortex that is transitional from limbic to association regions (reflecting the limited semantic content and immediate biological relevance of most calls). Asymmetries in the neural basis for call perception, and the existence of neurons in the auditory association cortex that selectively respond to conspecific calls, may be quite direct evolutionary precursors of the human neural substrate for language.

HUMAN LANGUAGE

Human language differs from animal communication systems in many ways. Functionally, as we have seen, most animal communication systems serve the purposes of identifying members of a species and designating items of immediate biological significance. Human language provides means to designate an infinitely large number of items, actions, and properties of items and actions. Human language also allows items, actions, and properties to be related to one another, and for relationships between events and states of affairs in the world, such as temporal order and causation, to be represented at the level of discourse. Language expresses these different semantic values with different types of representations: words, words made from other words, sentences, and discourse. Each of these "levels" of linguistic representation consists of specific forms that are related to specific aspects of meaning. The forms at each level are intricately structured, and there are an infinite number of different structures that can be built at each level of the system.

Language Processing

Current models of language processing subdivide functions such as reading, speaking, auditory comprehension, and writing into many different, semi-independent components, which are sometimes called "modules" or "processors." These components can be further divided into variable numbers of highly specialized operations, such as those involved in mapping features of the acoustic signal onto phonemes or in constructing syntactic structures from words. Each operation accepts only particular types of representations as input and produces only specific types of representations as output.

Models of language processing are often expressed as flow diagrams (or "functional architectures") that indicate the sequence of operations of the different components that perform a language-related task. Figure 49.2 presents a basic model illustrating the functional architecture of the lexical processing system. Figure 49.2 is greatly simplified because not all components are represented, parallel feedforward and feedback pathways are not depicted, and the nature of the operations in each component is not specified.

Some important operating characteristics of components of the language processing system emerged from experimental studies. These include the following features:

1. Language processors are highly specialized. For instance, recognition of phonemes is accomplished by mechanisms that appear to separate very early in the processing stream from those that recognize other auditory stimuli.
2. Language processors are obligatorily activated when an individual attends to the input. For instance, if we

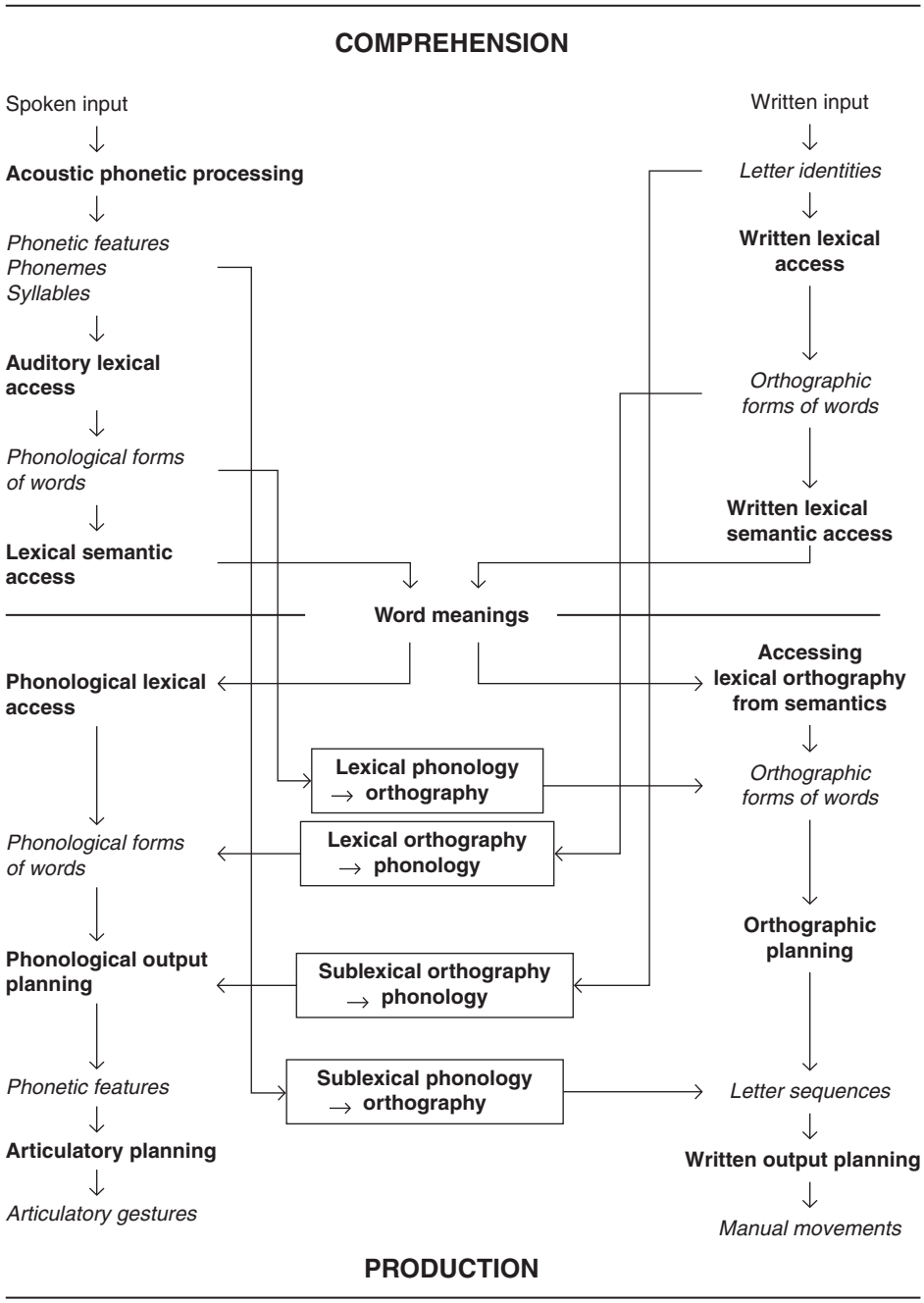


FIGURE 49.2 A model of the major psycholinguistic operations involved in processing simple words.

attend to a sound that happens to be the word "elephant," we must hear and understand that word; we cannot hear this sound as just a noise.

3. Language processors generally operate unconsciously.
4. Language processors operate quickly and accurately. For instance, it has been estimated on the basis of many different psycholinguistic experimental techniques that spoken words are usually recognized less than 125 ms after their onset—that is, while they are still being uttered. The speed of the language processing system as a whole occurs because of the speed of each of its components, but also is achieved because of the massively parallel functional architecture of the system, which leads to many components of the system being simultaneously active.
5. Language processors require "processing resources." Although we do not appreciate it consciously, language processing demands some effort. It is unclear whether there are separate pools of processing resources for each language processing

component or for language processing as a whole or whether language processing can "borrow" resources from other systems if it needs to.

Language develops, without explicit instruction, in the auditory-oral modality, if this channel is available. Prelingual deafness leads to language developing in the visual and gestural modalities. Languages can also be represented orthographically, using a variety of scripts (alphabetic, syllabic, consonantal, ideographic). Many of the features of spoken language processing we have reviewed apply to the processing of written and signed language as well, though there are differences due to constraints imposed by the modality of presentation. For example, written language does not develop in the absence of explicit instruction.

Language Universals, Language Acqusition, and Language Evolution

By some counts, there are over 6500 languages in the world. These languages differ in many ways from one another: they have different vocabularies, use different sounds, and have different ways of forming words and different syntactic rules. However, beyond these surface level differences, there are features that are common to all human language. For instance, the sound systems of all human languages consist of alternations of consonants and vowels; no language forms different words by varying features of sound such as its loudness or whether a word is whispered. These "linguistic universals" form a framework within which the features of individual languages occur.

Linguistic universals can arise because of universal features of motor or sensory systems or, possibly, because of constraints on cognitive computational capacities. They may also arise because language itself is constrained to have certain features. Chomsky's work in the area of syntax is the best known and most controversial example of analyses along these lines (Chomsky, 1965, 1995). Chomsky and his colleagues and students have argued that very abstract features of syntax are common to all languages. There have been numerous suggestions as to what these features are; one recent proposal is that they include the ability to generate an infinite number of expressions from a finite set of elements (Hauser, Chomsky, & Fitch, 2002). Many linguists believe that the structure of language is much less abstract than Chomsky and his colleagues have suggested and, accordingly, that universal "laws" of linguistic structure are also less abstract, mostly taking the form of statements that the presence of one type of element in a language implies that a second type will also be present.

The issues of how abstract linguistic representations are and what universal properties of language exist are bound up with two major issues—how language is acquired and how language evolved—that are in turn relevant to the neural basis for language. Chomsky's view is that language acquisition consists of fixing the parameters in an innate language template that correspond to the constraints on possible forms of human language that are present in the language a child is exposed to. A parallel in nonhuman learning might be imprinting (Immelmann, 1972), whose neurological basis has been studied very little. Models of language that see language structures as much less abstract argue for less elaborate innate knowledge and a greater role for learning in the development of language abilities (e.g., Seidenberg, 1997; Tomasello, 2003).

The same disagreements affect ideas about how language evolved. In Chomsky's model, a critical neural development is the emergence of structures and processes in the brain that support universal properties of language. Assuming recursion is one of these universal properties, the finding that starlings can discriminate a recursive from nonrecursive sequences of notes (Gentner, Fenn, Margoliash, & Nusbaum, 2006) implies that some of these neural structures are not unique to humans, but if recent studies showing that some species of monkeys do not recognize recursive structures are true of all higher primates, the evolutionary path for these features must have taken a number of divergent paths and reemerged in humans. The major alternative to the view that language evolution consisted of the emergence of neural structures capable of supporting abstract organizational features that are found in all human languages is that language rests upon an evolutionary change that supports much less complex and abstract representations and the complexity of human language is a cultural development, like architectural complexity.

One well-known model of this sort that has a neural basis has been developed by Rizzolati and Arbib (1998), who have argued that the neural structures that are critical for language are modifications of the mirror neurons found in monkeys that fire both when the monkey grasps particular objects and when the animal sees a person grasping those objects. In Arbib and Rizzolati's view, these neurons encode the relation of a number of actors to an action and an object, and thus lay the basis for neural mechanisms that allow an arbitrary number of substitutions of actors and objects in relation to the action described by a verb. We note that the fact that culture influences what structures a language has (Everett, 2005) does not resolve this debate, because any neurologically normal human infant would presumably learn the language of his or her environment.

There are very limited empirical data regarding the neural and genetic basis for the evolution of human language. Lieberman (2006) argues that the critical evolutionary change that made for human language was not specifically neural but consisted of the descent of the root of the tongue into the oropharynx, which allows humans

to produce the inventory of speech sounds of a language. In his view, two genetic effects—increases in the size of the hominid brain related to the ASPM gene and, more importantly, changes in the ability to exploit the increased channel capacity in humans by changes in the FOXP2 gene—underlie the use of the resulting increase in the vocal repertoire for communicative purposes. Deficits in the FOXP2 gene in a single family have been associated with a variety of language and articulation problems, and the FOXP2 gene has been referred to as a (or "the") "language gene" by some researchers, However, the deficits in affected family members are not specific to syntax or word formation, and, as Lieberman argues, it is more likely that this gene codes for neuronal features of subcortical gray nuclei, in particular the basal ganglia and, to a lesser extent, the cerebellum that make for more elaborate and efficient control of sequencing abilities than for cortical structures that represent and process language *per se*.

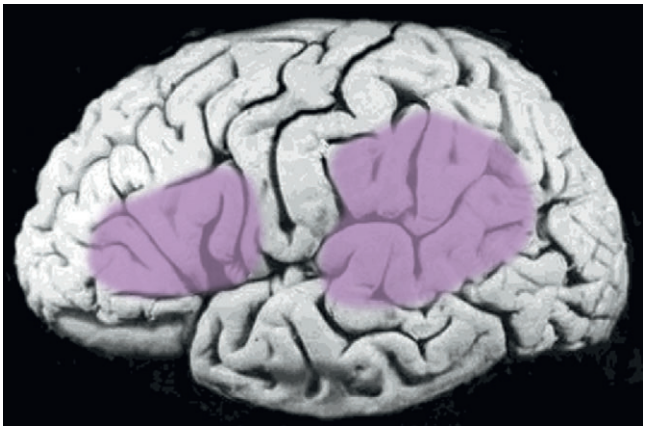


FIGURE 49.3 A depiction of the left hemisphere of the brain showing the main language areas. The area in the inferior frontal lobe is known as Broca's area, and the area in the superior temporal lobe is known as Wernicke's area, named after the nineteenth-century physicians who first described their roles in language. Broca's area is adjacent to the motor cortex and is involved in planning speech gestures. It also serves other language functions, such as assigning syntactic structure. Wernicke's area is adjacent to the primary auditory cortex and is involved in representing and recognizing the sound patterns of words.

NEURAL ORGANIZATION FOR LANGUAGE

An Inventory of Brain Structures Related to Language

Human language depends on the integrity of the unimodal and multimodal association cortex in the lateral portion of both cerebral hemispheres. This cortex surrounds the sylvian fissure and runs from the pars triangularis and opercularis of the inferior frontal gyrus [Brodman's areas (BA) 45, 44: Broca's area] through the angular and supramarginal gyri (BA 39 and 40) into the superior temporal gyrus (BA22: Wernicke's area) (Fig. 49.3). For the most part, the connections of these cortical areas are to one another and to the dorsolateral prefrontal cortex and lateral inferior temporal cortex (Fig. 49.3). These regions have only indirect connections to limbic structures. These areas consist of many different types of association cortex, devoted not to sensation or motor function but to a more abstract type of analysis. The nature of this cortex and its patterns of connectivity are thought to combine to give language its enormous representational power and to allow its use to transcend biological immediacy.

The evidence that language involves these cortical regions was originally derived from deficit-lesion correlations. Patients with lesions in parts of this cortex have been described who have had long-lasting impairments of language ("aphasia"). Disorders affecting language processing after perisylvian lesions have been described in many different types of disease, in all languages that have been studied, in patients of all ages and both sexes, and in both first and subsequent tongues, indicating that this cortical region is involved in language processing independent of these factors. Functional neuroimaging studies have documented increases in regional cerebral blood flow (rCBF) using PET or blood oxygenation level-dependent (BOLD) signal using fMRI in tasks associated with language processing in this region. Event related potentials whose sources are likely to be in this region have been described in relationship to a variety of language processing operations. Stimulation of this cortex by direct application of electrical current during neurosurgical procedures interrupts language processing. These data all converge on the conclusion that language processing is carried out in the perisylvian cortex.

Regions outside the perisylvian association cortex also appear to support language processing. Working outward from the perisylvian region, evidence shows that the modality of language use affects the location of the neural tissue that supports language, with written language involving the cortex closer to the visual areas of the brain and sign language involving brain regions closer to those involved in movements of the hands than movements of the oral cavity. Some ERP components related to processing improbable or ill-formed language are maximal over high parietal and central scalp electrodes, suggesting that these regions may be involved in language processing. Both lesion studies in stroke patients and functional neuroimaging studies suggest that the inferior and anterior temporal lobe is involved in representing the meanings of nouns. Activation studies also implicate the frontal lobe just in front of Broca's area in word meaning. Injury to the supplementary motor cortex along the medial surface of the frontal lobe can lead to speech initiation disturbances; this region may be important in activating the language

processing system, at least in production tasks. Activation studies have shown increased rCBF and BOLD signal in the cingulate gyrus in association with many language tasks. This activation, however, appears to be nonspecific, as it occurs in many other, nonlinguistic, tasks as well. It has been suggested that it is due to increased arousal and deployment of attention associated with more complex tasks.

Subcortical structures may also be involved in language processing. Several studies report aphasic disturbances following strokes in deep gray matter nuclei (the caudate, putamen, and parts of the thalamus). It has been suggested that subcortical structures involved in laying down procedural memories for motor functions, in particular, the basal ganglia, are involved in "rule-based" processing in language, such as regular aspects of word formation, as opposed to the long-term maintenance of information in memory, as occurs with simple words and irregularly formed words. The thalamus may play a role in processing the meanings of words. In general, subcortical lesions cause language impairments when the overlying cortex is abnormal (often the abnormality can be seen only with metabolic scanning techniques), and the degree of language impairment is better correlated with measures of cortical than subcortical hypometabolism. It may be that subcortical structures serve to activate a cortically based language processing system but do not themselves process language. The cerebellum has also increased its rCBF in some activation studies involving both language and other cognitive functions. This may be a result of the role of this part of the brain in processes involved in timing and temporal ordering of events, or in its being directly involved in language and other cognitive functions.

Lateralization

Most language processing goes on in one hemisphere, called the "dominant" hemisphere. Which hemisphere is dominant shows considerable individual differences and bears a systematic relationship to handedness. In about 98% of right-handed individuals, the left hemisphere is dominant. The extent to which left hemisphere lesions cause language disorders is influenced by the degree to which an individual is right handed and by the number of non-right-handers in his or her family. About 60 to 65% of non-right-handed individuals are left hemisphere dominant; about 15 to 20% are right hemisphere dominant; and the remainder appear to use both hemispheres for language processing. The relationship of dominance for language to handedness suggests a common determination of both, probably in large part genetic.

The neural basis for lateralization was first suggested by Geschwind and Levitsky (1968), who discovered that part of the language zone (the planum temporale—a portion of the superior temporal lobe; Fig. 49.4) was larger in the left than in the right hemisphere. Subsequent studies have confirmed this finding and identified specific cytoarchitectonically defined regions in this posterior language area that show this asymmetry. Several other asymmetries that may be related to lateralization have also been identified. The exact relationship between size and function is not known, as there are instances of individuals whose dominant hemisphere is not the one with the larger planum temporale. In general, however, relative size is a good predictor of lateralization. This is another example of the "bigger is better" principle that we saw applied to song nuclei in birds.

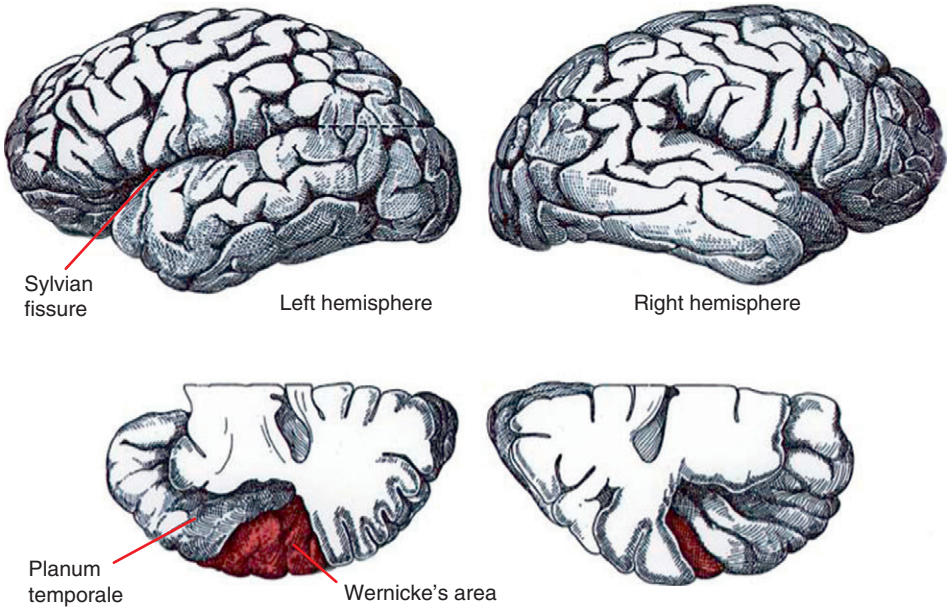


FIGURE 49.4 Depiction of a horizontal slice through the brain showing asymmetry in the size of the planum temporale related to lateralization of language

Although not as important in language functioning as the dominant hemisphere, the nondominant hemisphere is involved in many language operations. Evidence from the effects of lesions and split brain studies, experiments using presentation of stimuli to one or the other hemisphere in normal subjects, and activation studies all indicates that the nondominant hemisphere understands many words, especially concrete nouns, and suggests that it is involved in other aspects of language processing as well. Some language operations may be carried out primarily in the right hemisphere. The best candidates for these operations are ones that pertain to processing the discourse level of language, interpreting nonliteral language such as metaphors, and appreciating the tone of a discourse—for example, the fact that it is humorous. Some scientists have developed models of the sorts of processing that the right hemisphere carries out. For instance, it has been suggested that the right hemisphere codes information in a more general way compared to the left, representing the overall structure of a stimulus as opposed to its details. This may be a general feature of processing in the two hemispheres and apply to other functions, such as visual perception.

Organization of the Perisylvian Cortex for Language Processing

Historically, theories of the relationship of parts of the perisylvian association cortex to components of the language processing system have included distributed ("holist") and localizationist models. Contemporary studies have strongly supported the latter, and we will selectively review representative work.

Evidence for localization of language processing components comes from the finding that selective language deficits occur in patients with perisylvian lesions, often in complementary functional spheres. For instance, some patients have trouble producing the small grammatical function words of language (such as *the, what, is, he*), whereas others have trouble producing common nouns. The existence of these two disorders indicates that the tissue involved in producing function words is not involved in producing common nouns in the first set of patients, and vice versa in the second set.

The first localizationist theories were based on clinical observations in the mid- and late nineteenth century. The pioneers of aphasiology—Paul Broca, Karl Wernicke, John Hughlings Jackson, and other neurologists—described patients with lesions in the left inferior frontal lobe whose speech was hesitant and poorly articulated and other patients with lesions more posteriorly, in the superior temporal lobe, who had disturbances of comprehension and fluent speech with sound and word substitutions. These correlations led to the theory that language comprehension required the unimodal auditory association cortex (Wernicke's area, Brodman area 22) adjacent to the primary auditory cortex (Heschl's gyrus, Brodman areas 41) and that motor speech planning required the unimodal motor association cortex in Broca's area (Brodman areas 44 and 45) adjacent to the primary motor cortex (Brodman area 4). Geschwind (1965) added the hypothesis that word meaning was localized in the inferior parietal lobe (Brodman areas 39 and 40) because word meanings consist of associations between sounds and properties of objects, and the inferior parietal lobe is an area of multimodal association cortex to which fibers from the unimodal association cortex related to audition, vision, and somasthesis project.

This model has many limitations and problems. It only deals with words, not with other levels of the language code. The functions that are localized are entire language tasks, such as comprehension of spoken language or speech production, not components of the language processing system. The correlations between deficits and lesions are found primarily in the chronic stage of recovery from stroke, not in acute or subacute phases of stroke or in other diseases, and even in chronic strokes a very large number of lesions do not occur where the model predicts they will.

Modern studies describe deficits in aphasic patients in terms of specific linguistic representations and psycholinguistic processes. This is similar to modern studies using functional neuroimaging. In the next sections, we review studies of localization of two language processes: spoken word recognition and comprehension, and syntactic operations in comprehension. We have selected these functions because they contrast in several ways; they vary in their abstractness, the extent of the temporal intervals over which they integrate sensory input, and the extent to which current studies converge on localization of the operations they involve. They thus reflect the range of results found in current studies of localization of components of the language processing system.

Lexical Access and Word Meaning

Many of the processes involved in recognizing linguistically relevant units of sound, activating words from these units, and activating the meanings of words appear to rely on specific areas of the brain. This is not evident in the aphasia literature, which shows that both individuals with lesions involving the inferior frontal gyrus (IFG) and those with lesions involving the superior temporal gyrus (STG) display deficits in perceiving acoustic-phonetic stimuli linguistic categories such as phonemes (Blumstein, Cooper, Zurif, & Caramazza, 1977; Blumstein, Tartter, Nigro, & Statlender, 1984). However, results from functional neuroimaging studies with normal participants have provided evidence that speech perception recruits a neural processing stream involving both left posterior

and anterior brain structures (Binder & Price, 2001; Hickok & Poeppel, 2007; Scott & Wise, 2004) and that this neural stream comprises different processing stages that are located in specific areas.

The evidence suggests that temporo-spectral acoustic cues to feature identity are integrated in unimodal auditory association cortex lying along the superior temporal sulcus immediately adjacent to the primary auditory koniocortex (Binder, 2000). Some researchers have suggested that the unconscious, automatic activation of features and phonemes as a stage in word recognition under normal conditions occurs bilaterally. In their view, the dominant hemisphere supports controlled and conscious phonemic processing such as subvocal rehearsal, making explicit decisions in phoneme discrimination and identification and rhyme judgment tasks, and other similar functions.

Based on functional neuroimaging results, activation of the long-term representations of the sound patterns of words is thought to occur in the left superior temporal gyrus. Scott and her colleagues have argued that there is a pathway along this gyrus and the corresponding left superior temporal sulcus such that word recognition occurs in a region anterior and inferior to primary auditory cortex, and that word meanings are activated further along this pathway in anterior inferior temporal lobe bilaterally (Scott & Johnsrude, 2003; Scott & Wise, 2004). This pathway constitutes the auditory counterpart to the visual "what" pathway in the inferior occipital-temporal lobe.

Speech perception is connected to speech production, especially during language acquisition when imitation is crucial for the development of the child's sound inventory and lexicon. On the basis of lesions in patients with Reproduction Conduction aphasia, the neural substrate for this connection has been thought to consist of the arcuate fibers of the inferior longitudinal fasciculus, which connect auditory association cortex (Wernicke's area in the posterior part of the superior temporal gyrus) to motor association cortex (Broca's area in the posterior part of the inferior frontal gyrus). Recent functional neuroimaging studies and neural models have partially confirmed these ideas, providing evidence that integrated perceptual-motor processing of speech sounds and words makes use of a "dorsal" pathway separate from that involved in word recognition.

Hickok and Poeppel (2007) developed the first comprehensive statement of the two-pathway model (Fig. 49.5). They proposed that the ventral stream consists of an acoustic-phonetic conversion process located in STG bilaterally and a "lexico-semantic interface" in the posterior part of the left middle temporal gyrus, which is connected to inferior frontal gyrus. The ventral pathway (to middle temporal gyrus) is activated automatically and unconsciously in the process of ordinary spoken word comprehension. The dorsal pathway transmits phonetic representations from superior temporal gyrus to a left dominant temporo-parietal sensorimotor interface area (Area Spt) that interacts reciprocally with an articulatory network comprising the posterior inferior frontal gyrus, premotor area, and anterior insula. The dorsal pathway is activated in tasks that involve conscious and/or controlled meta-phonological processing and/or (possibly subvocal) rehearsal, such as phoneme discrimination, rhyme judgment, and tasks that involve verbal working memory.

As described above, traditional neurological models maintained that the meanings of words consist of sets of neural correlates of the physical properties that are associated with a heard word, all converging in the inferior parietal lobe. It is now known that most lesions in the inferior parietal lobe do not affect word meaning and functional neuroimaging studies designed to require word meaning do not tend to activate this region. Evidence is accruing that the associations of words include "retroactivation" of neural patterns back to unimodal motor and sensory association cortex (Damasio, 1989), and that different types of words activate different cortical regions. Verbs are more likely to activate frontal cortex, and nouns temporal cortex, possibly because verbs refer to actions and nouns refer to static items. More fine-grained relations between words and activation of sensory-motor brain areas have been described (Pulvermuller, 2005). Hauk, Johnsrude, and Pulvermüller (2004) found activation in primary motor and premotor cortex at the homuncular level of the leg when subjects recognize the written word "kick," the mouth when they recognize "lick," and the hand when they recognize "pick" (Fig. 49.6).

Both deficits and functional activation studies have suggested that there are unique neural loci for the representation of object categories such as tools (frontal association cortex and middle temporal lobe), animals and foods (inferior temporal lobe and superior temporal sulcus), and faces (fusiform gyrus) (see Caramazza and Mahon (2006) for review), and some of these localizations are adjacent to areas of the brain that support perception of features that are related to these categories (animals activate an area of lateral temporal adjacent to the area sensitive to biological motion). Debate continues about how these category effects arise, and whether these activations should be taken as reflecting the meanings of these words. At the same time as these specializations receive support, evidence from patients with semantic dementia and from functional neuroimaging indicates that another critical part of the semantic network that represents word meanings and concepts is located in the anterior inferior temporal lobes. It has been suggested that this region is a "convergence" area, where different properties of items are bound.

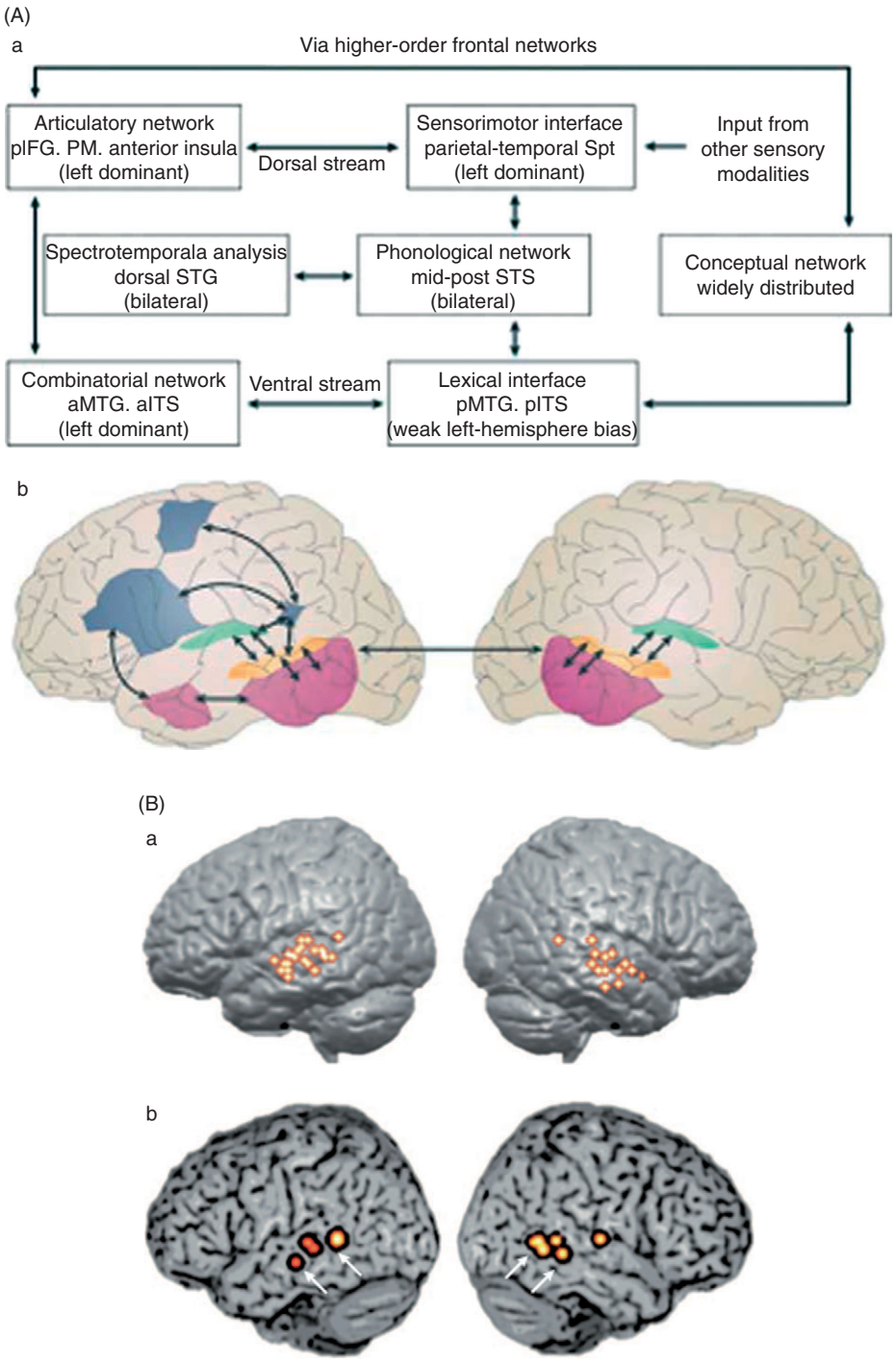


FIGURE 49.5 **(A)** Depiction of the lateral surface of the brain showing areas involved in the functional neuroanatomy of phonemic processing. H is Heschl's gyrus, the primary auditory cortex. STP is the superior temporal plane, divided into posterior and anterior areas. STG is the superior temporal gyrus. Traditional theories maintain that pSTP and STG are the loci of phonemic processing. Hickok and Poeppel (2000) argue that these areas in both hemispheres are involved in automatic phonemic processing in the process of word recognition. Other research suggests that more anterior structures, aSTP and the area around the superior temporal sulcus (STS), are involved in these processes. The inferior parietal lobe (AG, angular gyrus; SMG, supramarginal gyrus) and Broca's area (areas 44 and 45) are involved in conscious controlled phonological processes such as rehearsal and storage in short-term memory (from Hickok and Poeppel (2007)). **(B)** Activation of left superior temporal sulcus by the contrast between discrimination of phonemes and discrimination of matched nonlinguistic sounds (from Hickok and Poeppel (2007)).

Syntactic Operations

We mentioned above that one feature of language is that it allows words to be related to one another to express relations among items and actions and their properties. This "propositional" information is crucial to many human functions, including conveying information stored in semantic memory, planning activities, and others. Languages have "syntactic rules" that determine how words are related to one another. The existence of such rules allows language to express any relation among words, including unlikely or impossible relations, and therefore allows language to express hypothetical events and situations, which is critical to reasoning. Thus, though syntax is not a feature of language that is as familiar to people as words, it is a critical aspect of language. The neural basis of syntactically-based comprehension has been studied

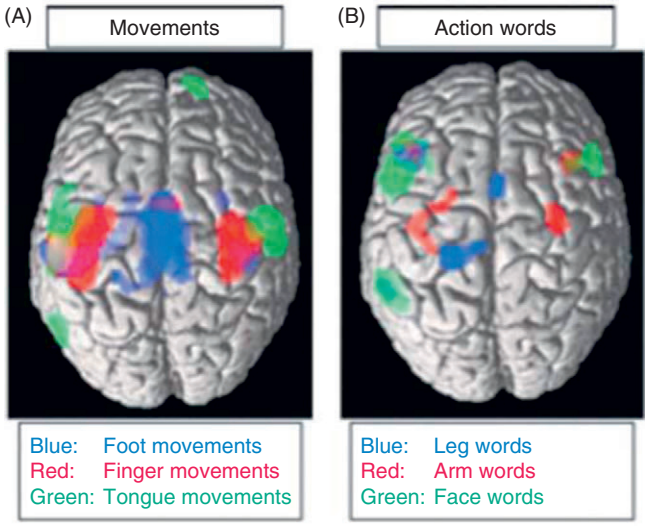


FIGURE 49.6 Activation of areas of the motor strip and premotor area by actions and action words (from Hauk et al. (2004)).

with both deficit-lesion correlations and functional neuroimaging.

Deficits of the ability to use syntax to understand sentences are typically established by showing that patients can understand sentences with simple syntactic structures (e.g., *The boy chased the girl*) and sentences with complex syntactic structures in which the relationships between nouns and verbs can be inferred from a knowledge of real-world events (e.g., *The apple the boy ate was red*), but not sentences with complex structures in which the relationships between nouns and verbs depend on a syntactic structure that needs to be constructed (e.g., *The boy the girl pushed was tall*). Patients with this pattern of comprehension disturbances were first described by Caramazza and Zurif (1976). Among the affected patients were ones with "Broca's aphasia," who produced "agrammatic" speech (speech that lacks function words and morphological markers). Caramazza and Zurif noted that the deficits in comprehension and expressive language of these patients both involved problems with function words, morphology, and syntax, and suggested that these deficits were connected. Their view was that the underlying deficit in these patients affected syntactic "algorithms," and suggested that these rules were supported by Broca's area (the left posterior inferior frontal gyrus), in which lesions were associated with expressive agrammatism. Variants of this hypothesis have been developed by a number of researchers, in particular Grodzinsky (2000), who linked lesions in this area to the loss of quite specific syntactic elements ("traces," in Chomsky's model of syntactic structure).

However, the data that are available indicate that deficits in syntactic processing in sentence comprehension occur in all aphasic syndromes and after lesions throughout the perisylvian cortex (Caplan, Baker, & Dehaut, 1985; Caplan, Hildebrandt, & Makris, 1996; Caramazza, Capitani, Rey, & Berndt, 2001). Conversely, patients of all types and with all lesion locations have been described with normal syntactic comprehension (Caplan et al., 1985). Caplan et al. (2007) studied 42 patients with aphasia secondary to left hemisphere strokes and 25 control subjects for the ability to assign and interpret three syntactic structures in enactment, sentence-picture matching, and grammaticality judgment tasks. In regression analyses, lesion measures in both perisylvian and nonperisylvian regions of interest predicted performance after factors such as age, time since stroke, and total lesion volume had been entered into the equations. Patients who performed at similar levels behaviorally had lesions of very different sizes, and patients with equivalent lesion sizes varied greatly in their level of performance.

Functional neuroimaging studies have also led several researchers to suggest that syntactic processing is localized in Broca's area or in portions of this region (Grodzinsky & Freiderici, 2006). However, most neuroimaging studies show that multiple cortical areas are activated in tasks that involve syntactic processing (Fig. 49.7) and, as with deficit-lesion correlation studies, different studies have shown different patterns of activation.

The discrepancies across studies using both patient data and activation techniques need to be understood. One possibility is that the neural tissue that is responsible for the operations underlying sentence comprehension and syntactic processing is localized in several neural regions, and that the extent to which an area is involved differs in different individuals. This would be consistent with the pattern of deficit-lesion correlations reported by Caplan et al. (2007) but is hard to reconcile with the neuroimaging results because, if it were true, one would expect there to be a variety of activation patterns in the individuals tested in any one study, most likely resulting in no specific activation in most studies. Another possibility is that different studies have involved different aspects of syntactic processing. This could be true of both patient and activation studies.

Pursuing this possibility, we note that poor performance in an aphasic patient can reflect many different deficits. For instance, chance level performance on a sentence type may result from a failure to assign the structure and meaning of a sentence, as many researchers have assumed, but could also result from an intermittent failure to assign structure and meaning. Recent studies using measures of on-line processing (self-paced word-by-word listening, and eye tracking) have provided evidence for the latter: some patients whose performance is at chance have normal on-line processing measures for sentences they interpret correctly and abnormal measures for sentences that they respond to erroneously. Examination of on-line performance in sentences that produced errors

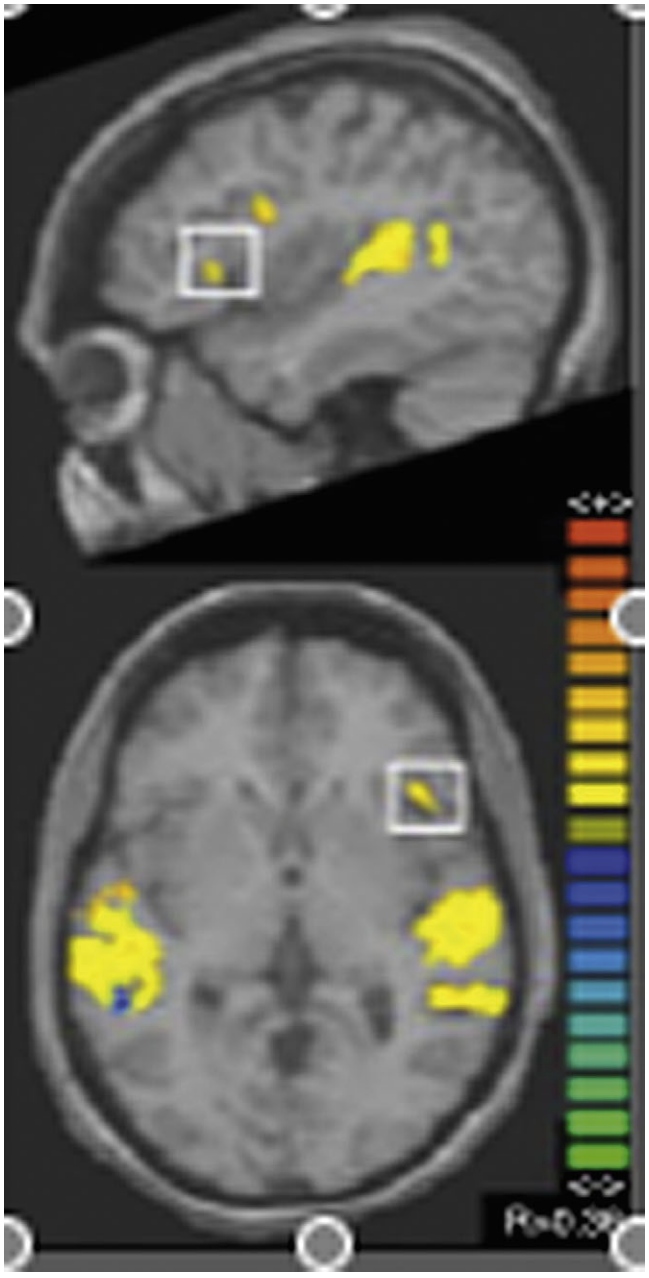


FIGURE 49.7 Activation of Broca's area and superior temporal sulcus when subjects processed sentences with particular syntactic features (questioned noun phrases) compared to matched sentences without those features *From Ben-Shachar, Palti, and Grodzinsky (2004).*

has suggested that a variety of deficits can cause comprehension failure. Some patients show abnormal on-line measures at points of greatest complexity (Caplan et al., 2007), suggesting intermittent overload of resource availability, and others have shown abnormal returns to previously relevant portions of a sentence after the point of maximum processing load has passed, suggesting that the level of a persisting competing interpretation was abnormally high. We have also found that syntactic comprehension deficits most often affect certain structures in only one task: for instance, a patient may be unable to match a sentence to one of two pictures but able to demonstrate its meaning by manipulating objects, suggesting that some deficits arise in the interaction of comprehension with satisfying the demands of different tasks. Different patterns of deficit-lesion correlation in different studies are consistent with localization of the processes that are affected in these different patients in different brain regions.

The variety of activation patterns seen in functional neuroimaging may have a similar explanation. Different sentence contrasts in different studies highlight different syntactic operations and may activate correspondingly different brain areas. In addition, activation related to performing a task, such as encoding sentence meaning in short term memory to answer a question, may also contribute to the results of many studies (Caplan, 2010), though much remains to be learned about the effects of these processes.

CONCLUSIONS

Human language is a unique representational system that relates aspects of meaning to many types of forms, each with its own complex structure. Animal communication systems are neither as complex nor as powerful as human language. Nonetheless, they have some similarities to language and provide clues as to the neural basis of human language. Some language traits are part of a shared evolutionary ancestry with other species, others represent convergent, independent evolution of analogous processes, while yet others must be unique to our species. Perhaps the most important lessons learned from animal studies are that complex communication systems can be innate, that learning can shape an innately specified range of behavioral possibilities, and that communication systems can rely on specific nuclei within the brain and be lateralized. Other features of the neural basis for animal communication systems, such as seasonal variation in the size of birdsong nuclei and the mechanisms that underlie this phenomenon, are not obviously relevant to human language.

Deficit-lesion correlations and neuroimaging studies are beginning to provide data about the neural structures involved in human language. It appears that one general area of the brain—the left perisylvian association cortex—is especially important in representing and processing language (although other regions are involved as well), and that, within this area, many language operations are localized in specific regions. Charting the exact areas that support language operations will require attention to both the nature of the language units that are activated (e.g., different types of sounds, different syntactic structures) and

the tasks that are being performed (e.g., unconscious automatic vs. controlled, conscious phoneme processing; encoding sentence meanings into memory vs. determining if a sentence is plausible). Many possibilities exist, such as that some language functions are localized in very specific regions and others are supported by more widely distributed areas of cortex or by multiple areas of cortex. The way the brain is organized to support different aspects of language still presents exciting challenges and opportunities for discovery for cognitive neuroscientists.

References

Ben-Shachar, M., Palti, D., & Grodzinsky, Y. (2004). The neural correlates of syntactic movement: Converging evidence from two fMRI experiments. *Neuroimage, 21,* 1320–1336.

Bickerton, D. (1990). *Language and species*. Chicago, IL: University of Chicago Press.

Binder, J. (2000). The new neuroanatomy of speech perception. *Brain, 123,* 2371–2372.

Binder, J. R., & Price, C. (2001). Functional neuroimaging of language. In R. Cabeza & A. Kingstone (Eds.), *Handbook of functional neuroimaging of cognition* (pp. 187–251). Cambridge, MA: MIT Press.

Blumstein, S. E., Cooper, W. E., Zurif, E. B., & Caramazza, A. (1977). The perception and production of voice-onset time in aphasia. *Neuropsychologia, 15,* 371–383.

Blumstein, S. E., Tartter, V. C., Nigro, G., & Statlender, S. (1984). Acoustic cues for the perception of place of articulation in aphasia. *Brain and Language, 22,* 128–149.

Caplan, D. (2010). Task effects on BOLD signal correlates of implicit syntactic processing. *Language and Cognitive Processes, 25,* 866–901.

Caplan, D., Baker, C., & Dehaut, F. (1985). Syntactic determinants of sentence comprehension in aphasia. *Cognition, 21,* 117–175.

Caplan, D., Hildebrandt, H., & Makris, N. (1996). Location of lesions in stroke patients with deficits in syntactic processing in sentence comprehension. *Brain, 119,* 993–949.

Caplan, D., Waters, G., Kennedy, D., Alpert, N., Makris, N., DeDe, G., … Reddy, A. (2007). A study of syntactic processing in aphasia II: Neurological aspects. *Brain and Language, 101,* 151–177.

Caramazza, A., & Mahon, B. Z. (2006). The organisation of conceptual knowledge in the brain: The future's past and some future directions. *Cognitive Neuropsychology, 23,* 13–38.

Caramazza, A., & Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and Language, 3,* 572–582.

Caramazza, A., Capitani, E., Rey, A., & Berndt, R. S. (2001). Agrammatic Broca's aphasia is not associated with a single pattern of comprehension performance. *Brain and Language, 76,* 158–184.

Cheney, D. L., & Seyfarth, R. M. (1990). *How monkeys see the world*. Chicago, IL: University of Chicago Press.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1995). *The minimalist program*. New York, NY: Praeger.

Damasio, A. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition, 33,* 25–62.

Everett, D. (2005). Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language. *Current Anthropology, 76,* 621–646.

Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature, 440,* 1204–1207.

Geschwind, N. (1965). Disconnection syndromes in animals and man. *Brain, 88,* 237–294, 585–644.

Geschwind, N., & Levitsky, W. (1968). Human brain: Left-right asymmetries in temporal speech region. *Science, 161,* 186–187.

Gould, J. L. (1982). *Ethology*. New York, NY: Norton.

Gould, J. L., & Gould, C. G. (1996). *Sexual selection* (2nd Ed.). New York, NY: Freeman.

Gould, J. L., & Gould, C. G. (1999). *The animal mind* (2nd Ed.). New York, NY: Freeman.

Gould, J. L., & Gould, C. G. (2007). *Animal architects*. New York, NY: Basic Books.

Gould, J. L., & Marler, P. (1987). The instinct to learn. *Scientific American, 256*(1), 74–85.

Grodzinsky, Y. (2000). The neurology of syntax: Language use without Broca's area. *Behavioral and Brain Sciences, 23,* 47–117.

Grodzinsky, Y., & Friederici, A. (2006). Neuroimaging of syntax and syntactic processing. *Current Opinion in Neurobiology, 16,* 240–246.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron, 41,* 301–307.

Hauser, M., Chomsky, N., & Fitch, W. T. (2002). The faculty of language. *Science, 298,* 1569–1579.

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences, 4,* 131–138.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews, 8,* 393–402.

Immelmann, K. (1972). Sexual and other long-term aspects of imprinting in birds and other species. *Advances in the Study of Behavior, 4,* 147–174.

Lieberman, P. (2006). *Towards an evolutionary biology of language*. Cambridge, MA: Belknap Press of Harvard University Press.

Pulvermuller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews, 6,* 576–582.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences, 21*(5), 188–194.

Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neuroscience, 26,* 100–107.

Scott, S. K., & Wise, R. J. S. (2004). The functional neuroanatomy of prelexical processing of speech. *Cognition, 92,* 13–45.

Seidenberg, M. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science, 275,* 1599–1603.

Tomasello, M. (2003). *Constructing a language: A usage based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Suggested Readings

Gould, J. L., & Gould, C. G. (1995). *The honey bee* (2nd Ed.). New York, NY: Freeman.

Grodzinsky, Y., & Amunts, K. (Eds.), (2006). *Broca's region*. Oxford, UK: Oxford University Press.

Hauser, M. (1996). *The evolution of communication*. Cambridge, MA: MIT Press.

Posner, M., & Raichle, M. E. (1997). *Images of mind*. New York, NY: W.H. Freeman.

Pulvermuller, H. (2002). *The neuroscience of language*. Cambridge, UK: Cambridge University Press.

David N. Caplan and James L. Gould

This page intentionally left blank

CHAPTER

50

The Prefrontal Cortex and Executive Brain Functions

INTRODUCTION

What controls your thoughts? How do you decide what to pay attention to? How do you act appropriately while dining in a restaurant or listening to a lecture? How do you manage to pursue long-term goals, like obtaining a college degree, in the face of the many distractions that can knock you "off task"? In short, how does the brain manage to orchestrate the activity of millions of neurons to produce behavior that is willful, coordinated, and extends over time? This is called cognitive control, the ability of our thoughts and actions to rise above mere reactions to the immediate environment and be proactive: to anticipate possible futures and coordinate and direct thought and action to them. It is a hallmark of intelligent behavior.

Because cognitive control necessarily involves the coordination of many different brain processes, it no doubt depends on circuits that extend over much of the cerebral cortex. However, one cortical region, the prefrontal cortex (PFC), seems to play a central role. It is the cortical area that reaches the greatest relative size in the human brain and is thus thought to be the neural instantiation of the mental qualities that we think of as "intelligent." Accordingly, the PFC receives converging information from many brain systems processing external and internal information, and it is interconnected with motor system structures needed for voluntary action. Here, we discuss anatomical, neuropsychological, and neurophysiological evidence for, and theories of, the role of the PFC in cognitive control.

ANATOMY AND ORGANIZATION OF THE PREFRONTAL CORTEX

The PFC consists of a collection of cortical areas that differ from one another on the basis of size, density, and distribution of their neurons. Its location in the human brain and major divisions in the monkey brain are shown in Figure 50.1, but these can be further subdivided into at least 18 distinct areas. The PFC is well positioned for a central role in cognitive control. Collectively, these areas have interconnections with brain areas processing external information (with all sensory systems and with cortical and subcortical motor system structures), as well as internal information (limbic and midbrain structures involved in affect, memory, and reward). The subdivisions have partly unique, but overlapping, patterns of connections with the rest of the brain, which suggests some regional differences in function (Fig. 50.2). As in much of the neocortex, however, there are local connections between different PFC areas that can result in an intermixing and synthesis of disparate information needed for cognitive control. Indeed, the PFC has long been suggested to be the great integrator, a brain region that synthesizes information about the external and internal world for the purpose of producing goal-directed behavior.

The PFC is Connected with Sensory Systems

The dorsolateral and ventrolateral portions of the PFC receive visual, auditory, and somatosensory information from the occipital, temporal, and parietal cortices, while

© 2013 Elsevier Inc. All rights reserved.

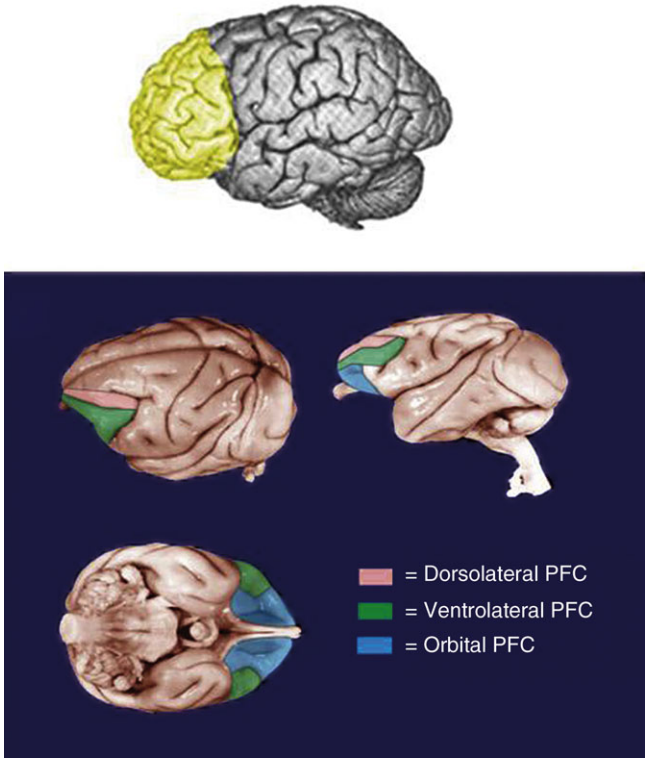


FIGURE 50.1 Location of the prefrontal cortex in the human brain (top) and its major subdivisions in the macaque monkey brain (bottom).

the orbitofrontal cortex receives chemosensory and visceral information (Fig. 50.2). Many PFC areas receive converging inputs from at least two sensory modalities. For example, the dorsolateral (areas 8, 9, and 46) and ventrolateral (12 and 45) areas both receive projections from visual, auditory, and somatosensory cortex. Further, the PFC is interconnected with other cortical regions that are themselves sites of multimodal convergence. Many PFC areas (9, 12, 46, and 45) receive inputs from the rostral superior temporal sulcus, which has neurons with bimodal or trimodal (visual, auditory, and somatosensory) responses to external stimulation. The arcuate sulcus region (areas 8 and 45) and area 12 seem to be particularly multimodal. They contain zones that receive overlapping inputs from all three sensory modalities. In all of these cases, the PFC is not directly connected with primary sensory areas, but instead interconnected with secondary or "association" sensory cortex.

The PFC is Connected with Motor System Structures

The dorsal PFC, particularly dorsolateral area 46, has preferential connections with many motor system structures and may be the primary region by which the PFC exerts control over behavior. Dorsolateral area 46 is interconnected with motor areas in the medial frontal lobe such as the supplementary motor area (SMA), pre-SMA, and the rostral cingulate, with the premotor cortex on the lateral frontal lobe, and with cerebellum and superior colliculus. Dorsolateral area 46 also sends projections to area 8, which contains the frontal eye fields, a region important for voluntary shifts of gaze. As with sensory cortex, there are no direct connections between the PFC and primary motor cortex; the PFC is instead connected with premotor areas that, in turn, send projections to primary motor cortex and the spinal cord. Also important are the dense interconnections between the PFC and basal ganglia (see following).

The PFC is Connected with the Basal Ganglia

The basal ganglia (BG) are a set of subcortical structures that are closely associated with both the motor system and the PFC. The anatomy of the BG is similar to that of the PFC in that they receive converging inputs from virtually every cortical system and many subcortical structures. The region of the BG that receives inputs, the striatum (which includes the caudate and putamen), receives a heavy projection from the PFC. Information then travels through loops of subcortical nuclei, which project back to the frontal cortex via the thalamus. While the PFC projects back to the same wide range of cortical areas that give input to the basal ganglia, the basal ganglia project largely to the frontal lobe motor system structures and the PFC. Thus, the BG are well positioned to influence a wide range of brain functions, albeit with an emphasis on motor output. The anatomical "loops" between the PFC and BG have inspired theories of their close interdependency. This will be discussed following. Basal ganglia anatomy and functions are discussed in greater detail in Chapter 30.

The PFC is Connected with the Limbic System

The orbital PFC is more closely associated with medial temporal limbic structures critical for long-term memory and "internal" information, such as affect and motivation, than other PFC regions. It has direct and indirect (via the thalamus) connections with the hippocampus and associated neocortex, the amygdala and hypothalamus. It has strong connections with olfactory and gustatory cortices, but weaker connections than the lateral PFC to parietal and temporal cortices.

Intrinsic PFC Connections May Provide a Substrate for Diverse Information to Interact

Like most of the neocortex, many PF connections are local. There are not only interconnections between the major subdivisions, but also interconnections between their constituent areas. The lateral PFC is particularly well

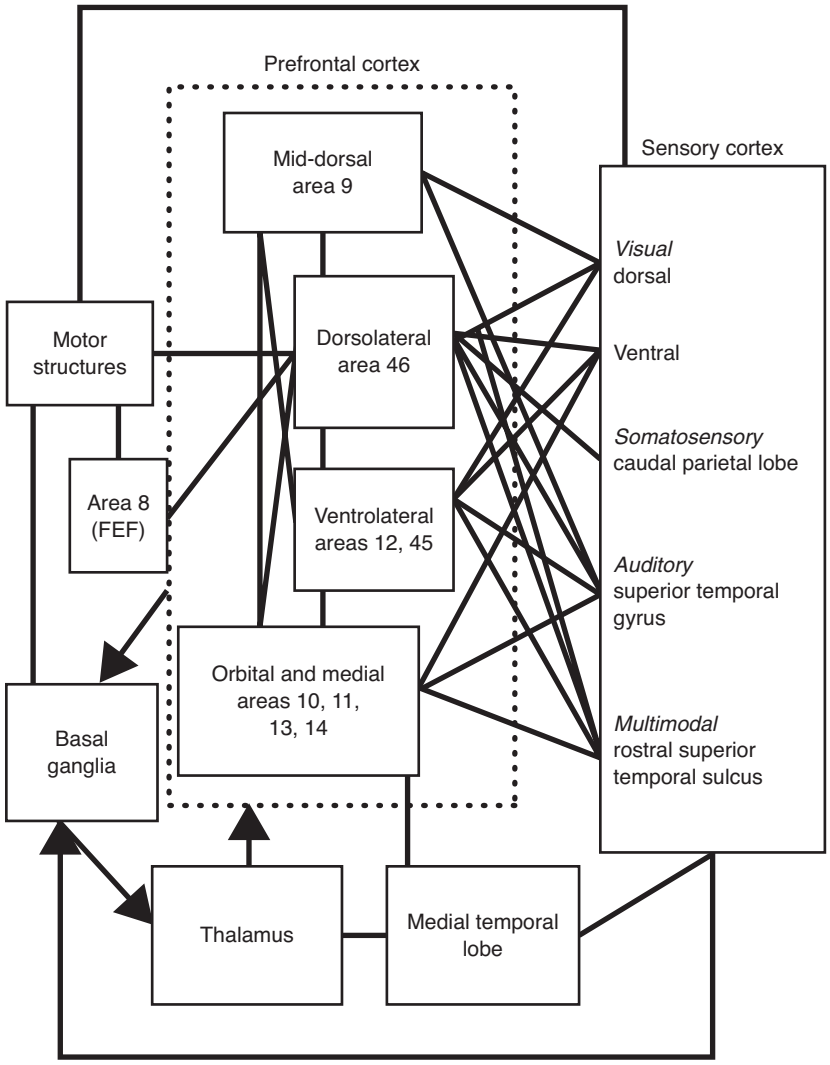


FIGURE 50.2 Schematic diagram of some of the extrinsic and intrinsic connections of the PFC. Most connections are reciprocal; the exceptions are noted by arrows. *From Miller and Cohen (2001).*

connected. Ventrolateral areas 12 and 45 are interconnected with dorsolateral areas 46 and 8, with dorsal area 9, and with orbitofrontal areas 11 and 13. Intrinsic PFC connections presumably allow information from a given PFC afferent or from a given process within a PFC subregion to be distributed to other parts of the PFC.

Summary

The PFC is anatomically well situated to play a role in cognitive control. It receives information from and sends projections to forebrain systems that process information about the external world, motor system structures that produce voluntary movement, systems that consolidate long-term memories, and systems that process information about affect and motivational state. This anatomy has long suggested that the PFC may be important for synthesizing external and internal information needed to produce complex behavior.

EFFECTS OF DAMAGE TO THE PREFRONTAL CORTEX IN HUMANS

PFC Damage in Humans Results in a "Dysexecutive Syndrome"

The PFC occupies a far greater proportion of the human cerebral cortex than in other animals, suggesting that it might contribute to those elusive cognitive capacities that separate humans from animals. However, unlike electrical stimulation of more posterior regions that produce hallucinations or motoric responses, stimulation of the PFC produces no obvious effect. This led it to be termed "silent cortex." Similarly, at first glance, PFC damage has remarkably little overt effect; patients can perceive and move, there is little impairment in their memory, and they can appear remarkably normal in casual conversation. The most infamous case is that of Phineas Gage, a railway construction foreman. In 1848, he suffered an accident in

which an iron bar exploded and was projected up through his chin and out of the top of his head, destroying much of his medial PFC in the process (Damasio, Grabowski, Frank, Galaburda, & Damasio, 1994) (Fig. 50.3). He survived, but it soon became clear that Gage had changed. While he had once been an efficient foreman, he was now irreverent and profane, rude to his coworkers, and unable to organize and plan the work that needed carrying out.

Despite the superficial appearance of normality, PFC damage seems to devastate a person's life. Take the case of Elliot, a successful, happily married accountant, and by all accounts a responsible person. In his late thirties, he was diagnosed as having a prefrontal meningioma, and during the removal of the tumor, his PFC was damaged. Within a few years of the operation, Elliot had divorced his wife, got married and divorced again, lost touch with his family and friends, acquired a disreputable business partner, and lost his business. All of Elliot's basic mental functions seemed to be intact, his language and memory abilities seemed normal, and on tests of intelligence he consistently performed in the superior range.

FIGURE 50.3 Reconstruction of damage to Phineas Gage's brain. *From Damasio et al. (1994).*

What Elliot seemed to lack was the ability to coordinate his mental functions in a manner that took long-term concerns (unseen goals) into account. He became impulsive, lurching from one catastrophe to another based on his momentary whims.

Note that this scenario is exactly what would be expected from the loss of a cognitive control system. Damage to the higher-order, executive functions leaves the components of complex behavior intact and able to be elicited by the appropriate external trigger. But the individual would not be able to coordinate these behaviors with respect to goals and task-specific constraints. For example, there was a PFC patient who, when preparing coffee, first stirred and then added cream. Because of this apparent dysfunction in the top-down control over behavior, the pattern of deficits following frontal lobe damage has been thought to reflect a lack of goal-directedness and the result is called the dysexecutive syndrome.

Humans with Frontal Damage Are Disinhibited

One of the most striking features of prefrontal patients is their disinhibition and lack of behavioral control. The patients are often impulsive, quick to anger, and prone to making rude and childish remarks. For example, one patient complained of being "much more outspoken" since her injury, while another stated, "I've become impulsive since the accident. If I have something to say, I can't wait and have to say it straight away."

An example of disinhibition is utilization behavior. During a medical examination, a patient with PFC damage will pick up and use items that have been left on the doctor's desk. For example, the doctor might leave a comb on the desk, and the patient will pick up the comb and start combing his hair. In one extreme case, a patient used a urinal that the doctor had left on the desk! More formal tests of disinhibition include the Stroop test (Fig. 50.4A). It requires the patient to name the color of the ink in which a word is written. The catch is that the word is the name of another color. Patients have difficulty inhibiting the saying of the word itself because the word elicits a strong, prepotent tendency to read it.

Such tests require patients to inhibit previously established specific responses, but PFC patients are also impaired at inhibiting more abstract behavioral demands like rules. Consider the Wisconsin Card Sorting Test (WCST) (Fig. 50.4B). The patient is required to sort a deck of cards on which are printed a number of symbols. The cards differ from one another in the color and shape of the symbols and the number of symbols per card. The patient's task is to sort the cards according to a given criterion—for example, "Sort by shape." At this point the patients perform the task as well as control subjects. However, when the criterion is changed—for example, to "Sort by number"—the patients are impaired; they are unable to

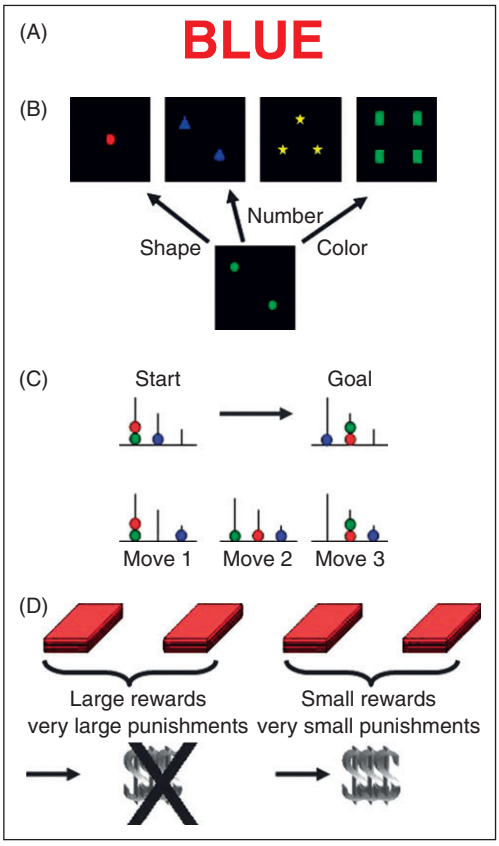


FIGURE 50.4 Tasks used to assess PFC dysfunction in humans. (A) Stroop task. (B) Wisconsin Card Sort. (C) Tower of London task. (D) Gambling task. See text for explanation of tasks.

inhibit sorting by the previous rule. This behavior that is continuing to perform a previously reinforced behavior even though it is no longer rewarded is called perseveration and is characteristic of a prefrontal patient's performance on a wide variety of tasks.

This inability to ignore irrelevant stimuli and the actions they trigger has led PFC patients to be described as stimulus-bound, and the disruption of the patient's ability to think and plan has been termed *goal neglect* (Duncan, Emslie, Williams, Johnson, & Freer, 1996). Without the influence of the goal-related signals from the executive functions, the systems always respond reflexively to an input with whatever output is strongly connected to it.

Humans with Frontal Lobe Damage Have Emotional Impairments

PFC damage produces a marked alteration of a patient's moods and emotions. In the case of Phineas Gage, these changes were so profound that his friends commented, "Gage is no longer Gage." Patients typically display an emotional shallowness, being indifferent and apathetic both to their own situation and to the needs of others. For example, one patient stated, "If I saw someone cry, I'd just laugh; people look silly getting upset." Formal tests show that patients are less able to recognize the emotional expressions of others, though their ability to identify faces remains intact. Patients also become more irritable, aggressive, and prone to angry outbursts. Their humor, termed *Witzelsucht*, becomes compulsive, puerile, and facetious. The patients are aware of these changes even if they are indifferent to them. For example, one patient commented, "Emotion, tears, that's all gone out of the window," while another stated, "I'm much more aggressive, and I feel much less fear. I go fighting for no reason; I don't get anxious."

Humans with Frontal Damage Have Difficulty in Planning and Decision Making

Prefrontal patients have a great deal of difficulty planning and organizing their lives. Wilder Penfield, a neurosurgeon during the mid-twentieth century, described the case of his sister, whom he operated on to remove a large frontal tumor. Prior to the surgery she had been an excellent cook, and while her basic culinary skills remained subsequent to the operation, she seemed unable to organize her behavior so that all the elements of the meal were ready simultaneously. Instead she would move haphazardly from preparing one part of the meal to another, so that some parts of the meal would be burned, while others had hardly been started.

Formal tests have been developed to measure the ability of patients to organize their behavior. For example, the Tower of London Test requires the patient to think several moves in advance (Fig. 50.4C). There are three vertical posts that have three colored rings that can be placed on the posts. The three posts are of different lengths such that the first post can accommodate all three rings, the second post can accommodate only two, and the third post just one. From an initial starting position the patient is asked to move the rings one at a time, from one post to another, to attain a specified goal position. This task requires the patient to plan a number of subgoal positions in order to reach the final goal. Patients with prefrontal damage are greatly impaired at this task, making many more moves than necessary to reach the final goal position. The Multiple Errands Test taxes more everyday planning abilities. The test takes place in a shopping center, and the patient is required to perform various errands, such as obtaining a copy of the previous day's newspaper. The patients take much longer than control subjects. They go into inappropriate (irrelevant) shops as they pass them and repeatedly become distracted from their principal goal.

Impairments in planning might be related partly to difficulties evaluating the consequences of actions. Faced with a number of different, competing actions, patients

might not choose wisely because they are less sure of the consequences of those actions. Evidence for this arises from a "gambling" task (Fig. 50.4D). The patient has to choose cards from four different decks in an effort to win play money from the experimenter. Each card wins the patient some money, but some cards also lose money, and then the patient has to pay the experimenter. Unbeknownst to the patient, cards from the first two decks will win large amounts of money but are also occasionally associated with very large losses, and repeatedly choosing from these two decks will result in a net loss. By contrast, cards in the other two decks win smaller amounts of money, but the losses are also a lot smaller, and, overall, choosing from these two decks will result in a net profit. Control subjects initially choose from the two decks associated with the larger rewards, but after a few very large losses, they switch to choosing from the other decks and earn a profit. Patients with PFC damage also initially choose from the decks associated with larger rewards but then fail to alter their behavior until they lose all their money. This task is appealing because it is closely related to the types of problems that prefrontal patients experience in real life.

It was initially thought that PFC was only required when relatively complex decisions were required, like in the gambling task described previously. However, more recent findings have shown that PFC patients, particularly those with damage to the orbital PFC, are impaired even on extremely simple decisions. For example, if patients are shown two colored swatches and asked to pick their favorite, they appear to choose at random (Fellows & Farah, 2007). They seem to have a problem using their subjective preferences to guide their decisions. However, they can still solve perceptual decisions—for example, if you show them two colored swatches and ask them which is more blue. Therefore, it is thought that orbital PFC may be particularly important for value-based decision making—in other words, assigning a value to two options and then making your decision based on that value (Rangel, Camerer, & Montague, 2008).

Humans with Frontal Lobe Damage Have Impaired Working Memory

Patients might also have difficulties planning due to a problem with holding information in mind for a short while. This is the kind of short-term memory that one relies on when reading a phone number and then going to dial that number. This function is called working memory because the information is forgotten as soon as it is no longer relevant. The concept of working memory is influenced by the observation that monkeys with lesions of the PFC have deficits on a spatial delayed-response task that requires them to remember the location of a stimulus or behavioral response over a brief delay of several seconds. Early studies failed to find this deficit in humans with prefrontal damage, possibly because the patients verbalized the location of the stimulus (e.g., "It's on the left"), which presumably makes the task much easier. Later studies, however, in which the prefrontal damage was more extensive and bilateral did show a spatial delayed-response deficit. Other studies used abstract patterns that could not be easily verbally encoded and found that the patient's nonspatial working memory was also impaired.

Summary

We have seen that the PFC poses a complex problem. Damage to a relatively circumscribed region of cortex in humans produces a bewildering array of symptoms. But one caution in interpreting clinical studies concerns the size of the damage that typifies the prefrontal patient. Not only does the damage often encompass many different PFC regions, each of which may have different functions, but it also often encroaches onto neighboring regions, such as the premotor cortex, the cingulate cortex, and the insular cortex. Plus, it should be remembered that the PFC comprises many different anatomically defined areas, each possessing unique patterns of connections with the rest of the brain; it might be that each component of the dysexecutive syndrome arises from damage to different areas of the PFC. These matters are difficult to explore in humans with PFC damage, because the damage is typically so extensive that it encompasses many brain regions both within and external to the PFC. There are two solutions to this problem. The first is to use neuroimaging techniques, such as functional magnetic resonance imaging (fMRI), in healthy humans. This technique can detect areas of increased neuronal activity in specific brain regions with a resolution of 2–4 mm^3. A second method is to study the effects of experimentally produced focal lesions in monkeys. We will discuss each of these approaches in turn.

NEUROIMAGING STUDIES AND PFC

Neuroimaging studies have shown that many of the tasks with which patients with PFC damage struggle also activate PFC in healthy subjects. However, an advantage of the technique is that specific regions within PFC can be identified that are activated preferentially by specific cognitive processes. For example, fMRI studies show that performing the WCST activates ventrolateral PFC. In general, a number of trends have emerged from this field regarding the functional organization of PFC.

The clearest functional distinction is between orbitofrontal cortex and lateral PFC. Orbitofrontal cortex is consistently activated by tasks requiring the processing of

rewarding or emotional information. For example, it shows more activation to pleasant images (e.g., a baby's face) or unpleasant images (e.g., mutilated bodies) as opposed to neutral images. Similar results are seen in other modalities such as pleasant or unpleasant smells, tastes, sounds, and touches. Orbitofrontal cortex even responds to abstract rewards such as winning or losing money. Orbitofrontal cortex is not simply responding to the presence of these rewards, but seems to use this information to guide our decisions and choices. For example, orbitofrontal cortex is activated when subjects are choosing a meal from a menu, but not if they are simply reading the menu with no intention of ordering.

In contrast, lateral PFC is activated by a wide assortment of cognitive tasks. These include tests of inhibitory control (such as the Stroop task), rule switching tasks (such as the WCST), and working memory tasks. A distinction that has emerged from these studies concerns the respective roles of dorsolateral and ventrolateral PFC regions. Ventrolateral PFC is predominately activated by tasks that require holding information in working memory while dorsolateral PFC is activated when that information must be manipulated in some way. We return to this idea later in the chapter when we discuss the "multiple processing levels" theory of prefrontal function. Recently investigators have begun to focus on the most anterior portions of PFC, referred to as the "frontal pole," which corresponds to area 10. There is some evidence that this region might be responsible for controlling the transfer of information between dorsolateral and ventrolateral PFC.

In sum, neuroimaging studies have helped to show that we cannot consider the PFC as a single unitary area but that distinct subregions of PFC are responsible for carrying out specific processes. The clearest distinction is between the emotional and reward-based processing occurring in orbitofrontal cortex and the more cognitive processes occurring in lateral PFC. A problem with fMRI, however, is that while it can identify which regions are more active during a given task, it does not tell us whether those regions are *necessary* for the performance of the task. To determine whether a given PFC subregion is necessary for a specific cognitive process, findings from monkey neuropsychology are useful. We discuss these findings in the next section.

EFFECTS OF DAMAGE TO THE PREFRONTAL CORTEX IN MONKEYS

Studying the effects of experimentally induced lesions in monkeys has two advantages. First, because the lesions are experimentally induced, they can be small and focused on specific subregions of PFC. Second, they are a strong test of whether a specific subregion is necessary for the performance of a given cognitive task.

In 1935, Jacobsen published a seminal paper showing that lesions of the frontal lobe in monkeys produced deficits on a spatial delayed-response task. In this task, the monkey sees a reward hidden at one of two locations, and then after a brief delay is allowed to retrieve it (Fig. 50.5A). Monkeys with large PFC lesions behave as if they forgot where the reward was hidden, even after delays of just a few seconds. Subsequent experiments showed that lesions restricted to the dorsolateral PFC alone could produce the same deficit. By contrast, damage to the ventrolateral PFC impaired object delayed-response tasks. Monkeys with such damage have difficulty with delayed matching-to-sample (DMS) tasks in which they are shown an object (the sample) and then, after a brief delay of several seconds, must choose the same object (the match) from one or more nonmatching objects (Fig. 50.5B). The most severe deficits are observed when the same two objects are used repeatedly, requiring the monkey to remember which of the two objects it has seen most recently as the sample. Monkeys with orbital (ventral) PFC lesions tend to show the greatest deficit in tasks requiring reversal of previously successful actions. For example, they are impaired at spatial reversal tasks, in which the reward is alternated between two locations, (Fig. 50.5C) and object reversal tasks, in which a reward is alternated between each of two objects (Fig. 50.5D).

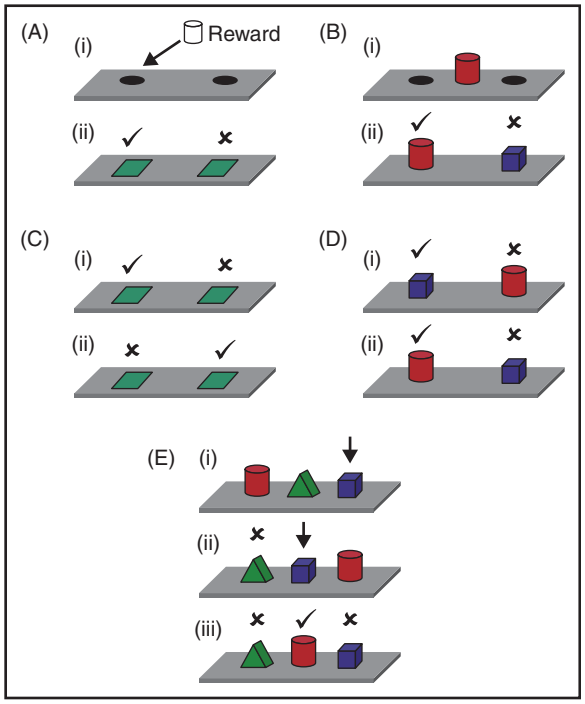


FIGURE 50.5 Some of the tasks on which monkeys with PFC damage are impaired. Check indicates correct choice, ✗ indicates incorrect choice. Arrow in E indicates monkey's choice in self-ordered task. (A) Spatial delayed response. (B) Delayed matching-to-sample. (C) Spatial reversal. (D) Object reversal. (E) Object self-ordered search task.

In both tasks the monkeys perseverate, continuing to choose the location or object under which the reward had first been hidden. Such observations led to the suggestion that the dorsal PFC might be responsible for processing spatial information, the ventral PFC for object-related information, and the orbital PFC for inhibitory control in accord with their preferential connections with the parietal, inferior temporal, and limbic system structures, respectively.

The picture, however, is not so simple and straightforward; a clear double dissociation between these impairments has not been demonstrated. Consider the hypothesized spatial versus object dissociation in the PFC. Dorsolateral PFC can impair performance on tasks that do not have a spatial component, such as object self-ordered search tasks. In this task the monkey is presented with three objects covering rewards. They must move each object in turn to get the reward, but after each choice, a screen is lowered and the object's positions are shuffled (Fig. 50.5E). Thus, the monkey must continually remember which of the three objects it has already checked for a reward. Neurophysiological studies paint a similar picture. While there may be some bias in the distribution of neurons that are sensitive to object identity or location in space (especially for the small population of highly specialized "face cells"), there is also a good deal of intermixing of neurons encoding these attributes, especially when the integration of these attributes is relevant to task performance. For more discussion of the issue of PFC functional topography, see Box 50.1.

Monkeys with PFC Damage Are Impaired on Rule-Based Tasks

We have already seen that humans with PFC damage have difficulties on tasks requiring the patient to use rules, such as the WCST (Fig. 50.4B). Marmosets were trained on a task similar to the WCST. They had to choose complex stimuli on the basis of a rule, such as "Choose by shape." When they were required to switch between different rules, monkeys with lesions of the dorsolateral PFC perseverated by continuing to choose on the basis of the previous rule, while monkeys with lesions of the orbital PFC were unimpaired. In contrast, when the same stimuli were used in an object reversal task, monkeys with lesions of the orbital PFC perseverated by continuing to choose the stimulus that had initially been rewarded, but the performance of monkeys with lesions of the dorsolateral PFC was normal. This suggests that both the dorsolateral and orbital PFC contribute to inhibitory control, but the dorsolateral PFC may be more engaged by inhibitory control over rules, while the orbital PFC may be more engaged by inhibitory control over choices based on objects and rewards.

Monkeys with frontal lobe damage also show deficits on simpler rule tasks, such as conditional learning. This requires the animal to solve a conditional problem of the form "If X, then Y"—for example, "If you see a red light, then move the lever to the right." Conditional learning requires associative relationships to be learned that are

BOX 50.1

FUNCTIONAL TOPOGRAPHY

Information about topography of neural properties in the PFC can provide important insights into its function. For example, knowing that primary visual cortex is organized into groups of neurons with similar tuning for orientation provides important clues to how it analyzes visual inputs. The anatomically complex nature of the PFC, coupled with the unique patterns of connections that each prefrontal area possesses, would certainly suggest a degree of functional fractionation. The clearest anatomical dissociation that has emerged is between the orbital and medial PFC regions and the lateral and dorsal PFC regions. Orbital and medial PFC areas are implicated in "hot" processes, such as motivation and emotion. Lateral and dorsal PFC areas implement "cool" cognitive processes, such as working memory and reasoning.

However, although there is this broad functional segregation, a more fine-grained functional topography has yet to be realized. One immediate problem that investigators face when trying to address this issue is that it is not clear how to subdivide cognitive processing. It is far easier to specify the sensory attributes of an apple than it is to list the different processes involved in planning or reasoning. In searching for this more fine-grained topography we should bear in mind two important points. First, PFC functional topography is likely to be based on broad categories or gradients of organization with overlap and intermixing of diverse information. Complex behavior requires that we recognize and respond to relationships across diverse dimensions, and representing these relationships is a prerequisite for cognitive control. Indeed, the majority of outputs of a given prefrontal area are to other prefrontal areas, so it is not surprising that many prefrontal neurons are multimodal. Second, because learning is central to cognitive control and modulates the properties of many PFC neurons (see Neurophysiology of the Prefrontal Cortex), it should have a large influence on PFC topography. Indeed, the flexible nature of controlled behaviors likely requires a dynamic anatomy, with neuronal properties changing to reflect the demands of the current task.

arbitrary and extend beyond the simple one-to-one mappings that underlie reflexive reactions. Whether or not a given response is successful depends on additional, contextual information. For example, reaching for popcorn can be rewarding, but only if one takes other information into account; if the popcorn belongs to another patron, the result could be disastrous. The need to take into account complex relationships in order to decide among alternative actions is presumably why cognitive control has evolved.

The advantage of using monkeys, however, is that the location and extent of the damage are under the control of the experimenter. This allows for more precise localization of the cortex necessary for a given process. This revealed that learning simple conditional rules seemed to depend on the premotor cortex immediately posterior to PFC, whereas learning more high-level abstract rules, of the kind tested by the WCST, involved dorsolateral PFC. Badre and colleagues showed that these findings extended to humans. Using fMRI, they showed that there was a posterior-anterior gradient in PFC, with progressively more abstract rules represented in progressively more anterior parts of PFC (Badre & D'Esposito, 2007). Conditional rules activated premotor cortex, rules based on stimulus dimensions (e.g., color or shape) activated dorsolateral prefrontal cortex, while rules based on the temporal context in which the task occurred activated the most anterior part of PFC, the frontal pole.

Monkeys with PFC Damage Show Emotional Impairments

Like humans, monkeys with PFC lesions, particularly in the orbitofrontal region, display marked changes in social and emotional behavior. Monkey society is dictated by a dominance hierarchy, and each monkey knows its place; a monkey is submissive to monkeys higher in the hierarchy (for example, the monkey grooms them) and is aggressive and threatening to monkeys lower in the hierarchy. Monkeys with PFC lesions, however, disregard this hierarchy and also decrease their interactions with troop members. This typically results in their being ostracized from the troop.

Current theories suggest that orbitofrontal cortex plays an important role in processing the value of sensory stimuli. For example, suppose that you learn that pressing the red button on a vending machine delivers chocolate, whereas pressing the green button delivers chips. Suppose also that you have just eaten a big bar of chocolate. The red button is now less valuable to you than the green button (since you are sick of chocolate!), and you are more likely to select the green button. Indeed, this is exactly how monkeys trained on a similar task behave. However, monkeys with damage to orbitofrontal cortex make choices more randomly, and their choices no longer appear to reflect the current value of the stimuli.

This inability to respond appropriately depending on the value of sensory stimuli could well underlie the social deficits that animals with orbitofrontal damage exhibit. Effective social behavior requires a constant monitoring of socially relevant stimuli. For example, paying attention to a person's facial expressions while you are speaking to them can yield all sorts of clues that you may need to modify what you are saying. They could look confused, upset, or angry. A failure to adjust to these cues would then be manifest as socially inappropriate behavior.

Recent results have also specified differences in the function of different parts of orbitofrontal cortex. Lateral orbitofrontal cortex is responsible for determining which outcome went with which stimulus, while the medial orbitofrontal cortex is responsible for comparing options and deciding which you would prefer (Walton, Behrens, Buckley, Rudebeck, & Rushworth, 2010). In the previous vending machine example, lateral orbitofrontal cortex would be responsible for storing the information about which button produces which food, while the medial orbitofrontal would be responsible for deciding which food you currently prefer. Together they can control which button you press. The authors also performed sophisticated behavioral analyses to show that dysfunction of these processes could produce perseverative deficits on reversal tasks (Figs. 50.5C, D). In other words, orbitofrontal cortex has little to do with inhibitory control. Rather the problems with reversal tasks that follow orbitofrontal damage reflect a more general problem in tracking which outcome was associated with which sensory stimulus.

Summary

Humans with damage to the PFC present a wide variety of symptoms, including a lack of inhibitory control, a shallowness of emotion, difficulty planning and organizing behavior, and impaired working memory. In general, they seem to have difficulty in exerting volitional control and organizing their mental processes and actions in order to direct them toward unseen goals. As a result, their behaviors are simple, reactive, and emitted without concern for future consequences.

Just like humans, monkeys with extensive prefrontal damage can appear remarkably unimpaired to the casual observer. They have no motoric impairments, and their ability to discriminate a variety of sensory stimuli remains intact. They are also capable of simple learning, such as Pavlovian and instrumental conditioning. However, they are impaired when behavioral demands extend beyond this to include additional task-specific constraints and/or involve overcoming a previously established response. Localizing a specific deficit to any one PFC subregion has proven difficult; double dissociations between different PFC regions are rare.

The pattern of deficits observed after PFC damage in humans and monkeys makes sense in light of the cognitive architecture depicted in Figure 50.2. The deficits seem to reflect a selective loss of higher-level functions. Lower-level, well-established routines are intact, but without the information about future goals and means provided by the higher-level functions, processing would be at the mercy of the environment. Behavior would be ruled by whatever sensory inputs happen to flow into the system and by whatever thoughts, emotions, and actions are strongly associated with these inputs. Behavior would seem impulsive and disinhibited, and emotional reactions inappropriate, because they would be emitted reflexively without any consideration of the future. Further, without the influence of future goals to continually drive task-appropriate processes, individuals would be distractible and would easily go "offtrack" when there are temporal gaps between sensory inputs, or between those inputs and the individual's responses.

NEUROPHYSIOLOGY OF THE PREFRONTAL CORTEX

PFC anatomy suggests it has the means for executive functions, and the effects on behavior after damage to the PFC support this idea. So do neurophysiological studies; they have shown that PFC neurons have the properties one would expect from a region intimately involved in goal-directed behavior.

The PFC Is Multimodal

PFC anatomy indicates that it is interconnected with many different brain systems, a prerequisite for cognitive control. This high degree of connectivity is reflected in the neural activity observed in the PFC. Its neurons are activated by visual, auditory, tactile, and gustatory stimulation (and their memory), as well as voluntary limb and eye movements. Further, these neurons seem capable of synthesizing this information in accord with task demands. For example, after monkeys are trained to make arm movements to visual and auditory targets appearing in different locations, many lateral PFC neurons are activated by both kinds of targets. Interestingly, few bimodal neurons were found in studies in which monkeys passively experienced visual and auditory stimuli. This finding suggests that task demands can exert a strong modulatory influence on PFC activity, which is consistent with its putative role in cognitive control.

PFC Neurons Can Sustain Their Activity

The first observations of PFC neural activity that were related to a cognitive process were made in the 1970s by Fuster, Niki, and their colleagues (Kubota & Niki, 1971; Fuster, 1973). Guided by lesion studies in monkeys, they recorded from the PFC while monkeys performed delayed-response tasks. They found that, during the memory delay, many PFC neurons showed sustained activity that seemed to keep information about a spatial or object cue online (Fig. 50.6). In spatial delayed-response tasks, cue-related "delay activity" could, in principle, reflect either sensory information related to the cue or movement information related to the behavioral response. Both are found in the PFC, each with different temporal characteristics. Sensory-related activity tends to be stronger nearer in time to the cue and then weakens as time passes, whereas movement-related neurons tend to increase their activity over time as the movement draws nearer. Goldman-Rakic (1987) and colleagues used a refined oculomotor version of the delayed-response task to demonstrate that dorsolateral PFC neurons have precise tuning for memories of particular visual field locations (Funahashi, Bruce, & Goldman-Rakic, 1989) (Fig. 50.7). They also demonstrated that PFC delay activity depends on dopamine receptors, a neurotransmitter implicated in signaling reinforcements and thus central to acquiring and sustaining goal-directed behaviors.

PFC delay activity has been shown to convey a wide range of behaviorally relevant cues: objects, colors, sounds, the frequency of a vibration to the hand, and forthcoming eye and arm movements. Similar activity is also evident in the sensory cortical areas and motor areas primarily responsible for analyzing inputs and generating behaviors, especially the higher-order areas that are

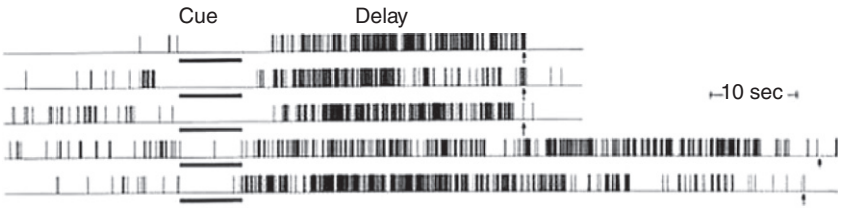


FIGURE 50.6 PFC delay activity. This figure shows the activity of a single PFC neuron during five trials of a spatial delayed response task. The arrow indicates the monkey's behavioral response at the end of the memory delay. Each small vertical line indicates an action potential from the neuron. Note the increased activity in the delay relative to other epochs. *From Fuster (1973).*

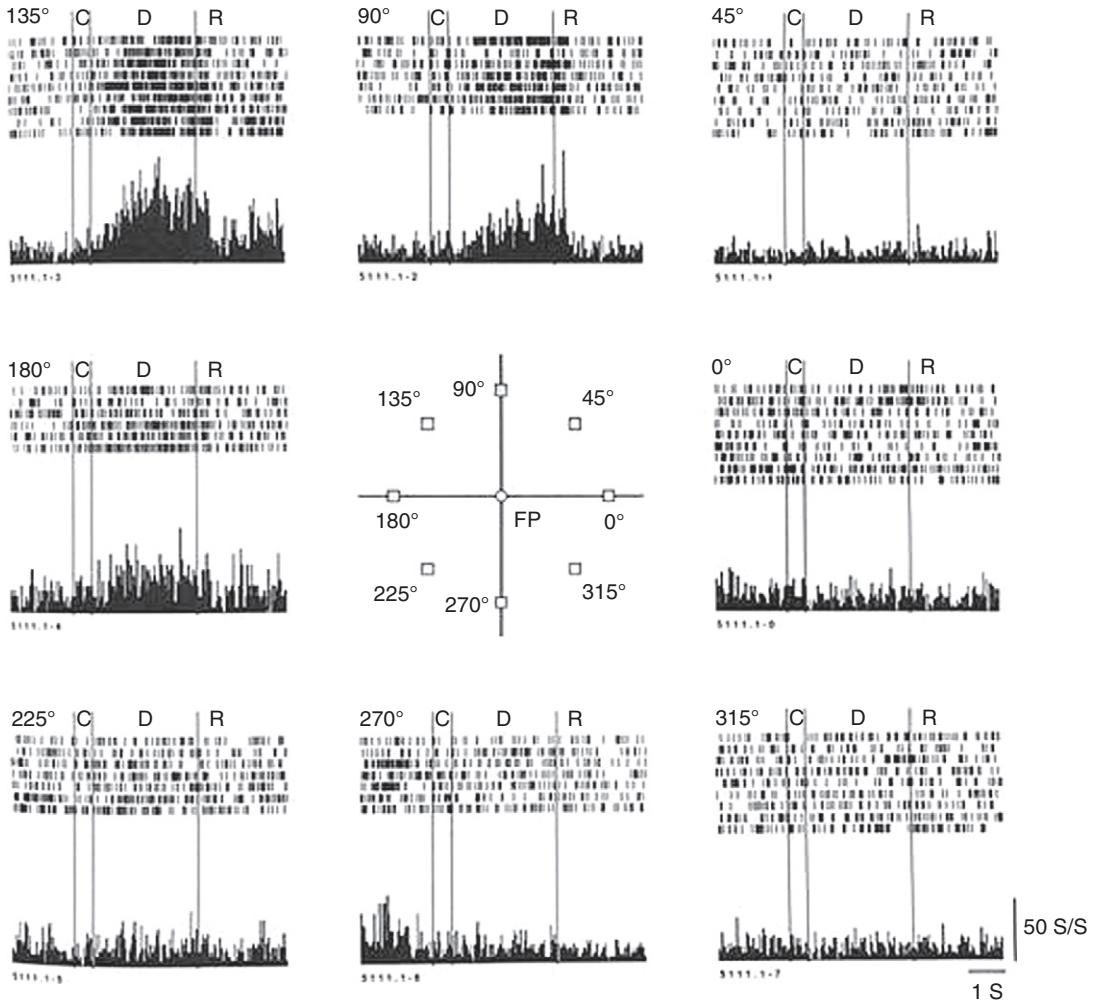


FIGURE 50.7 Activity of a single PFC neuron during performance of an oculomotor spatial delayed response task. Each trial starts with the monkey fixating the central fixation point (FP) indicated in the central diagram. After a delay (D), the monkey was allowed to respond by moving its eyes to the remembered location of the cue. Neural activity from cueing each location is shown by the corresponding histogram. The scale bars in the lower right indicate one second of time (1 S) and 50 spikes per second of neural activity (50 S/S). *From Funahashi et al. (1989).*

one or more steps removed from primary sensory and motor cortex. This is not surprising; sustained activity is apparent in many brain structures and must play a role in many neural processes. For example, it is likely to underlie iconic memory, the very brief lingering "afterimage" of recent sensory events. However, what separates more "cognitive" working memory processes from such lower-level processes as iconic memory is their ability to robustly sustain memories over potential distractions. PFC neurons do have this ability. For example, when monkeys are required to sustain the memory of a sample object across a delay period filled with visual distractors that require attention and processing, sustained activity within the prefrontal cortex can still maintain a memory of the sample object. By contrast, sustained activity in visual cortical areas seems more labile; it is disrupted by the presence of distractors. Following a distracting stimulus, neural activity reflects the distractor rather than the stimulus the monkey is retaining in working memory. This particular ability is not unique to the PFC. Some neurons in the entorhinal cortex, a region critical for memory, can also maintain sample-specific delay activity across intervening stimuli.

How does the PFC "latch" onto goal-relevant information and maintain it without disruption? Several models have been suggested. These typically employ a form of gating signal that instructs the PFC network when to maintain a given activity state. This gating signal may come from dopaminergic (DA) neurons in the midbrain. These neurons respond when monkeys are rewarded or, importantly, when monkeys can first predict that a reward is forthcoming. In many behavioral tasks, presentation of the cue at the start of the trial is a good predictor that reward is soon coming. Such a well-timed DA burst could strengthen current PFC representations, protecting them against interference

from disruption by irrelevant, distracting information until another DA influx reinforces another representation. Local injection of DA antagonists in the PFC does block delay activity in the PFC. The anatomical loops between the PFC and basal ganglia have been suggested to play a role here. Neural activity propagating around the loops may help sustain activity, and the basal ganglia have been suggested to provide a means by which DA signals can influence PFC activity. Box 50.2 describes the role of DA and PFC in Attention Deficit Hyperactivity Disorder.

PFC Neural Activity Reflects Task Rules

Playing the game requires knowing the rules. To explore how these rules might be represented in the PFC, experimenters manipulate task demands: monkeys are trained on tasks in which the physical cues are held constant while their behavioral meaning changes.

These studies suggest that PFC neurons represent task contingencies, the associative relationships that describe the logic needed to perform a task. Further support comes

BOX 50.2

ATTENTION DEFICIT HYPERACTIVITY DISORDER

Do you pay attention to captivating yet irrelevant stimuli? Do you have trouble acting appropriately, fidgeting while listening to a lecture? Do you have difficulty planning and pursuing long-term goals? Do distractions knock you "off task"? When these not uncommon difficulties occur to a severe degree, they are symptoms of Attention Deficit Hyperactivity Disorder (ADHD), a disorder seen in children and adults. It is estimated that 3 to 5% of childen have this disorder, with about half retaining problems into adulthood. The symptoms of ADHD involve dysfunction of the prefrontal cortex (PFC) and its cortical and subcortical connections.

The criteria for diagnosis of ADHD include symptoms of "inattention" and "hyperactivity/impulsivity." Patients can have either the combined type, or predominantly "inattentive" or predominantly "hyperactive/impulsive" types. Many of the symptoms for "inattention" used to diagnose ADHD relate to attentional abilities of the PFC—for example, difficulty sustaining attention or organizing, easily distracted and forgetful. Similarly, many of the symptoms of "hyperactivity/impulsivity" describe PFC deficits—for example, difficulty awaiting turn. The PFC controls attention via its projections to the parietal and temporal cortices (see Chapter 49), while it controls motor responses via its projections to the motor cortices and striatum. These circuits appear to be impaired in patients with ADHD. ADHD symptoms are especially evident in "boring" settings that require endogenous rather than exogenous regulation of behavior. For example, ADHD children can sit still and play video games for hours, but have trouble attending in school. ADHD patients have difficulty sustaining a behavior or thought over a delay. As described in this chapter, PFC cells exhibit sustained firing over a delay which regulates thought and behavioral output.

Structural and functional imaging studies show consistent alterations in PFC-striatal-cerebellar circuits in ADHD patients. Volumetric measures have detected smaller right-sided PFC regions (the right side is most associated with attention, and the right inferior prefrontal cortex with behavioral inhibition; see Chapter 49). Changes in the size of the striatum also have been reported, and functional imaging studies have shown abnormal activity of both the PFC and striatum in ADHD patients performing tasks that require PFC inhibitory or attentional functions. Structural imaging studies also have shown consistent decreases in the size of the cerebellar vermis, a region that may exert regulatory influences on the noradrenergic (NE) cells of the locus coeruleus (LC) and the dopaminergic (DA) cells of the ventral tegmental area. These NE and DA cells in turn have profound influences on PFC-striatal circuitry (see following). Thus, a smaller vermis in ADHD patients may lead to impaired catecholamine inputs to PFC and striatum. Some evidence supports this idea: a neuroreceptor imaging study suggests there are reduced numbers of catecholamine terminals in the PFC of adults with ADHD.

NE and DA have a critical influence on PFC-striatal circuits, and thus changes in these inputs can have tremendous import on PFC cognitive functions. The current chapter describes how DA has an important effect on PFC function, and the same is true for NE. NE cells fire when a stimulus is of interest to the animal, releasing NE in the PFC. NE stimulates post-synaptic α_{2A}-adrenoceptors, enhancing delay-related firing and strengthening regulation of behavior and attention. This is accomplished by α_{2A}-adrenoceptor inhibition of cAMP production, resulting in the closure of HCN channels (Hyperpolarization-activated Cyclic Nucleotide-gated cation channels). HCN channels are co-localized with α_{2A}-adrenoceptors on the dendritic spines of PFC neurons in the superficial layers of PFC, especially on spine necks. Thus, α_{2A}-adrenoceptors are ideally localized to control the neurochemical microenvironment around these channels. When cAMP builds up, HCN channels open, and synaptic inputs onto the spines are effectively shunted, thus functionally disconnecting PFC neuronal networks. In contrast, when α_{2A}-adrenoceptor stimulation inhibits the production of cAMP, HCN channels are closed and the networks are able

BOX 50.2 (*cont'd*)

to connect to control attention and behavior. In contrast, very high levels of NE and DA, such as released during stressful conditions, can impair PFC function via excessive cAMP production (e.g., from very high levels of β1 and DA D1 receptor stimulation), and from activation of high levels of protein kinase C signaling initiated by NE α_1-adrenoceptor stimulation. Thus, the PFC has to have just the right neurochemical conditions to function optimally.

What causes ADHD? ADHD is highly heritable and has been linked with genetic alterations that interfere with NE or DA signaling. These linkages include the DA transporter; the D1, D5, and D4 receptors (the D4 receptor also has very high affinity for NE); the synthetic enzyme for NE (dopamine beta hydroxylase); and the α_{2A}-adrenoceptor. Environmental factors may also lead to disruption of PFC-striatal circuits. For example, some kinds of infection may lead to autoantibodies that attack PFC-striatal circuits, and perinatal injury may impact these circuits.

Medications for ADHD likely reduce symptoms of inattention and impulsivity by optimizing the neurochemical environment in the PFC and in the striatum. All effective treatments for ADHD interact with catecholamines: Stimulants such as Ritalin (methylphenidate) and Adderall (a mixture of amphetamines) increase NE and DA release and/or block catecholamine reuptake. Recent research in rats suggests that low, therapeutic doses of these compounds preferentially increase release in PFC and have a greater effect on the release of NE than DA. Consistent with these data, NE reuptake blockers are also used to treat ADHD (e.g., desipramine), including highly selective NE agents (atomoxetine, also known as Strattera). Drugs that mimic NE at α_{2A}-adrenoceptors (e.g., guanfacine) are effective, especially in decreasing impulsivity and distractibility. Wellbutrin (also known as Zyban or bupropion) is a DA reuptake blocker that is used to treat ADHD. How do these medications actually work to ameliorate ADHD symptoms? There have been several ideas proposed: (1) the etiology of ADHD involves insufficient catecholamine innervation of PFC-striatal circuits, and drugs that increase catecholamine receptor stimulation correct this; (2) ADHD involves excessive catecholamine release, and low doses of these medications correct this by promoting negative feedback, decreasing phasic catecholamine release; or (3) catecholamines are normal in ADHD patients (e.g., it is simply that their cortex is immature or inflamed), and amplification of catecholamine mechanisms enhances PFC function and reduces symptoms. Support for this latter idea arises from the fact that normal individuals are improved by small doses of these medications as well, such as when college students sometimes take low doses of stimulants to focus on writing a paper.

All of us experience PFC dysfunction as part of daily life, especially when we are tired or stressed. These effects are likely due to neurochemical changes—for example, in NE and DA—that have great impact on the functional integrity of the PFC. But for people with ADHD, dysfunction of the PFC is much more constant and often beyond their control. Perhaps it is this prevalence of PFC dysfunction in otherwise healthy individuals, and its association with stress, that lead to the dismissal of ADHD by some as not a real disorder. It is hoped that as we come to better understand the neurobiological bases of ADHD, we will gain a fuller perspective of ADHD and its symptoms.

Amy Arnsten

Further Reading

Arnsten, A. F. T. (2006). Stimulants: Therapeutic actions in ADHD. *Neuropsychopharmacology, 31,* 2376–2383.

Arnsten, A. F. T. & Li, B. -M. (2005). Neurobiology of executive functions: Catecholamine influences on prefrontal cortical function. *Biological Psychiatry, 57,* 1377–1384. epub, Nov 17, 2004.

Solanto, M., Arnsten, A., & Castellanos, F. X. (Eds.). (2000). *The neuropharmacology of stimulant drugs: Implications for AD/HD*. New York: Oxford University Press.

Wang, M., Brian, P., Ramos, B. P., Paspalas, C. D., Shu, Y., Simen, A., … Arnsten, A. F. T. (2007). α2A-Adrenoceptors strengthen working memory networks by inhibiting cAMP-HCN channel signaling in prefrontal cortex. *Cell*. in press.

from investigations that have shown that PFC activity reflects more complex task information such as rules. Monkeys can be taught to rapidly alternate between different rules. For example, following a given cue, they can learn to direct a response to either the cue's location (spatial matching rule) or an alternate location associated with the cue (associative rule), depending on which rule is currently in effect. When tested in this fashion, many PFC neurons show rule-specific activity (Fig. 50.8). For example, a PFC neuron might respond to a given visual cue when the monkey is using an associative rule but exhibit weak or no activity under identical sensory and attentional conditions that differed only in that the monkey was instead using a spatial rule. Such effects have been found in PFC for associative versus spatial rules, for object matching versus spatial matching versus associative rules, and for shape matching versus object matching rules (Asaad, Rainer, & Miller, 2000).

Primate behavior is not limited to following literal associations and specific rules. Primates can also abstract

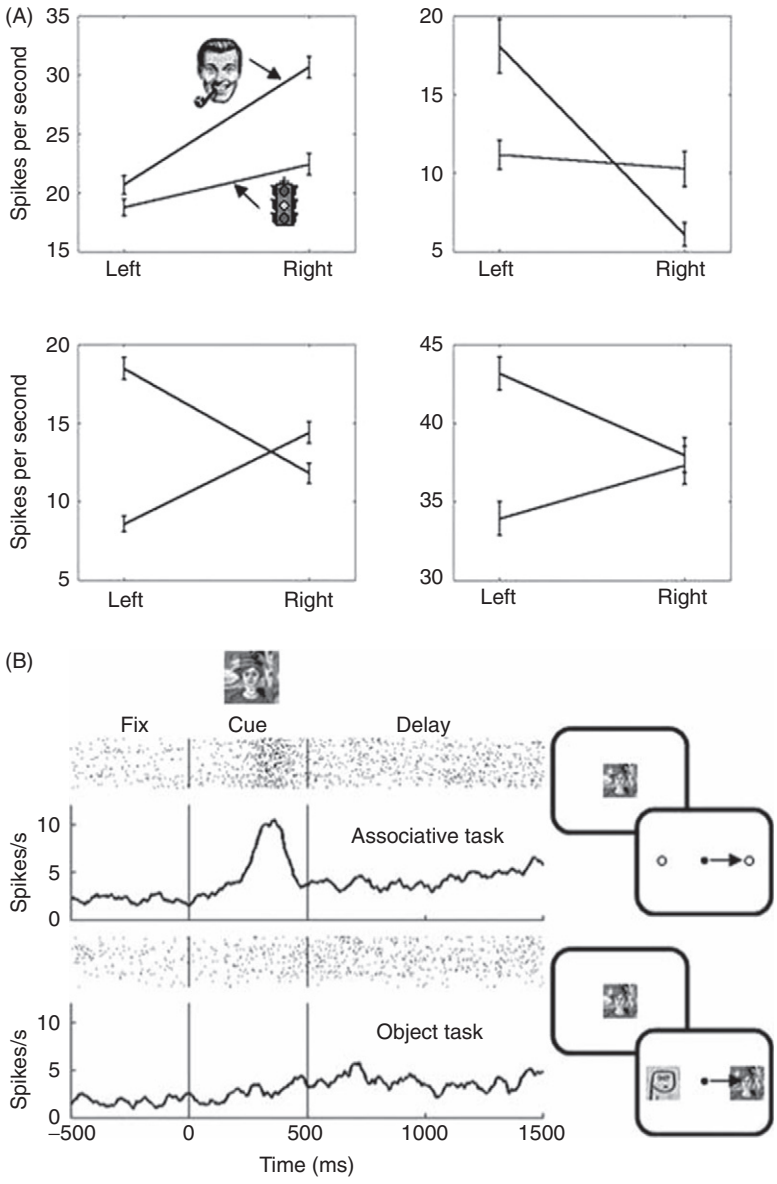


FIGURE 50.8 (A) The activity of four single PFC neurons when each of two objects, on different trials, instructed either a saccade to the right or a saccade to the left. The lines connect the average values obtained when a given object cued on the other saccade. The error bars show the standard error of the mean. Note that in each case, the neuron's activity depends on both the cue object and the saccade direction and that the tuning is nonlinear or conjunctive. That is, the level of activity to a given combination of object and saccade cannot be predicted from the neuron's response to the other combinations. (*Adapted from Asaad, Rainer, and Miller (1998)*). (B) A PFC neuron whose neural response to a cue object was highly dependent on the rule the monkey is currently using. The bottom half shows an example of a single PFC neuron's response to the same cue object during an object task (delayed matching to sample) and during an associative task (conditional visual motor). Note that the neuron is responsive to the cue during one task but not during the other, even though sensory stimulation is identical across the tasks. *Adapted from Asaad et al. (2000).*

general rules or principles, not tied to any particular stimulus or response. This ability allows behavior to extend beyond specific circumstances that have been directly experienced. For example, humans learn the "rules" for restaurant dining from specific experiences and can then generalize and apply them to new restaurants. Wallis, Anderson, and Miller (2001) investigated this process in monkeys by recording from single neurons in the PFC. Monkeys were trained to use two abstract rules; they indicated whether two successively presented pictures were either the same ("match") or different ("nonmatch"). The monkeys could perform this task with novel pictures, thus demonstrating that they had learned two general principles that could be applied to stimuli that they had not directly experienced. The most prevalent neuronal activity observed in the dorsolateral, ventrolateral, and orbital

PFC reflected the coding of the abstract "match" and "nonmatch" rules. This finding suggests that rule representation may be a cardinal PFC function. Recent findings using fMRI showed that humans also activate PFC when performing the exact same task.

PFC Neurons are Flexible

Another hallmark of primates is the ability to quickly learn new behaviors. PFC neurons are capable of a correspondingly rapid modification of their properties to meet task demands. For example, Cromer, Roy, and Miller (2010) trained monkeys to recognize two different and independent categorical distinctions (Cats vs. Dogs, Sports Cars vs. Sedans). They recorded neural activity from the PFC while monkeys switched back and forth between these categories. This revealed that many PFC neurons reflected *both* categorical distinctions. That is, the PFC neurons were category generalists that represent different, unrelated categories. This adaptability allows the brain to reutilize the same pool of neurons for different tasks. Further, this ability of PFC to multitask different types of information may be key to its role in rule learning. Without it, the PFC would quickly run out of capacity to learn new rules (Rigotti, Rubin, Wang, & Fusi, 2010). Neural network models suggest that having a large population of PFC neurons with mixed selectivity to diverse stimuli solves this capacity problem (Rigotti et al., 2010). Interestingly, neurons in the lateral intraparietal area (LIP) in posterior parietal cortex also seem to have this multitasking property (Fitzgerald et al., 2011), although it is not yet clear if those properties are generated with LIP or received from the PFC in a top-down fashion (see "Top-Down Control of Posterior Cortex by the PFC").

PFC Neurons Encode Rewards

Feedback about the consequences of actions is key to acquiring new goal-directed behaviors. Consequently, a large proportion of PFC neurons encode information about rewards, particularly in the orbitofrontal region. For example, some neurons are activated by the delivery of a reward, while others respond when an expected reward is not delivered. PFC neurons also encode the type of reward that the monkey expects to receive, for example a piece of cabbage versus a raisin, as well as the quantity of reward the monkey expects.

Recent evidence suggests that orbitofrontal neurons are in fact encoding a precise measure of the value of the reward. Determining this required overcoming an obvious hurdle: how does the experimenter determine how much an animal values a reward? One approach is through barter. If I want three chickens for my pig, then we can say that I find pigs three times more valuable than chickens. By bartering with a monkey using different types of juice, researchers showed that many orbitofrontal neurons encoded the value of one juice reward relative to another (Padoa-Schioppa, 2006). Orbitofrontal neurons also encode many other parameters that make one outcome more or less preferable. For example, they encode the delay until a reward is received, or the risk, probability or confidence that a reward will be received (Fig. 50.9). This has led to the suggestion that PFo may have a specialized role in encoding attributes that directly regulate the value of an outcome (Wallis, 2007). Consistent with this idea, PFo heavily connects with limbic areas responsible for processing reward information, such as the amygdala, nucleus accumbens, and hypothalamus, as well as gustatory and olfactory cortices, placing it in an excellent anatomical position to calculate the value of rewards.

Summary

As might be expected from a brain region suspected to orchestrate complex behavior, neural activity in the PFC conveys information about a wide range of external and internal information. But its activity is not limited to simple responses to sensory events or actions. Rather, the behavioral context has a pervasive influence on PFC activity. After training, many (up to half) dorsolateral, ventrolateral, and orbital PFC neurons encode information about the formal demands of behavior: contingencies between a wide range of stimuli, actions, and reward consequences, and behavior-guiding rules. By robustly sustaining their activity, PFC neurons can keep this and other behaviorally relevant information online to guide behaviors that extend over time. Also consistent with its role in guiding goal-directed behavior, PFC neurons process information about the expected results of achieving goals: rewards.

THEORIES OF PREFRONTAL CORTEX FUNCTION

Working Memory

One proposal is that the cardinal function of the PFC is to hold recent sensory inputs and movement-related information in active short-term memory, or working memory (Goldman-Rakic, 1987). In this view, the PFC acts as a temporary scratch pad that sustains the inputs it receives from other cortical areas. This hypothesis has its roots in observations of deficits in performance of delayed-response tasks following PFC damage and of PFC neurons that show sustained activity over a memory delay. Other deficits following PFC damage are thought to be sequelae of the working memory deficit. For example, patients might be impaired at planning because they are unable to keep in mind all the different variables that need to be considered. This in turn might make them appear disinhibited as

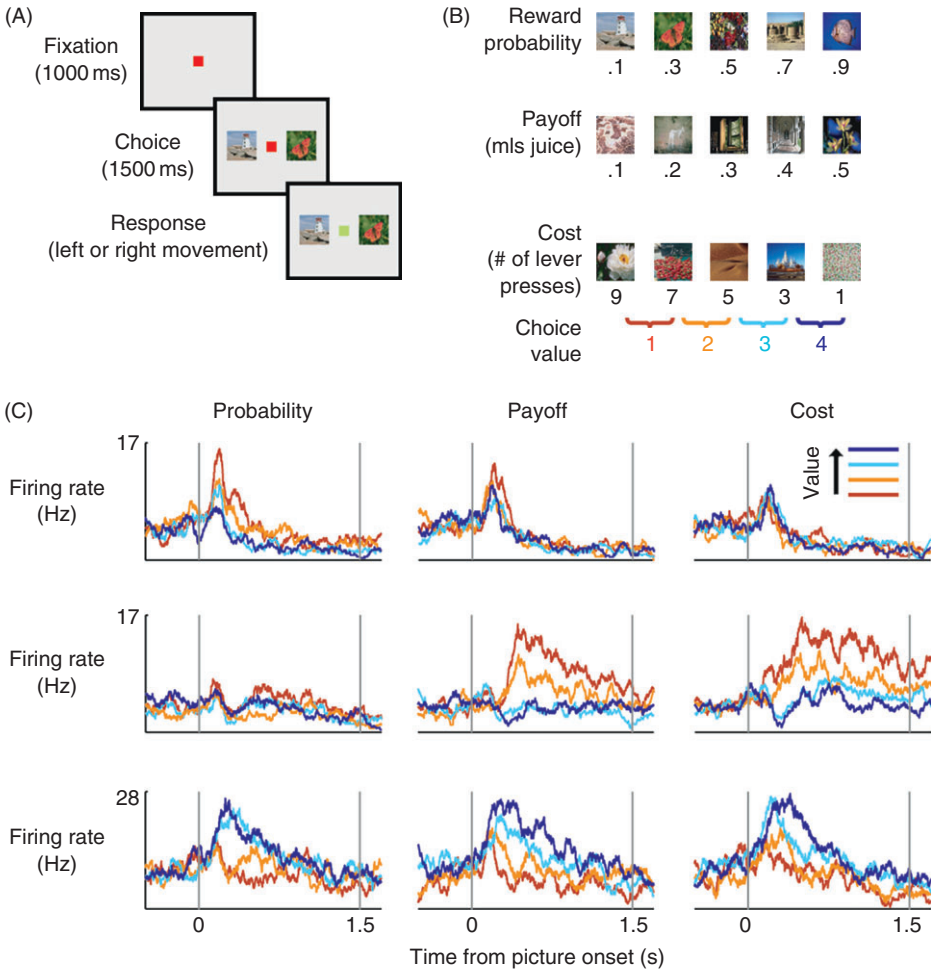


FIGURE 50.9 PFC neurons encode lots of information related to decisions. (A) The monkey was presented with two pictures and had to choose one. Each picture was associated with a specific outcome. (B) There were three sets of pictures. The probability picture set varied in terms of the likelihood of receiving a juice reward. The payoff picture set varied in terms of how much juice was associated with each picture. The cost picture set varied in terms of the number of lever presses necessary to earn the juice. Choice value indicates the choices that were used in the plots of neural activity. (C) Three neurons that encode different types of decision-related information. The top neuron encodes probability information. The second neuron encodes payoff and cost information. The bottom neuron encodes abstract value information that is independent of the precise type of decision. Some neurons encode increasing value, whereas other neurons encode decreasing value. *Adapted from Kennerely et al. (2009) with permission.*

their behavior becomes directed towards more immediate goals, rather than towards longer-term, more delayed goals.

Working memory maintenance functions are central to cognition; complex, goal-directed behaviors are typically drawn out and thus are impossible without mechanisms that extend information over time. There is ample evidence that these mechanisms are abundant in the PFC. Working memory is unlikely to be the whole story, however. Many of the tasks on which prefrontal patients and monkeys show impairment do not have an obvious working memory component, such as the WCST and reversal tasks.

Recent studies have explicitly investigated this issue, by examining the performance of monkeys with lesions of the ventrolateral PFC on a simultaneous color-matching task. In this task the monkey is presented with a sample color and two test colors simultaneously. The monkey must choose the test color that matches the sample color. PFC lesioned monkeys were impaired at this task, despite the fact the task involved contains no delays and therefore no working memory requirements. The task does, however, require conditional rules (e.g., "If the sample color is red, then choose the red test stimulus"), and the role of the PFC in the representation of such rules (see following) might explain this impairment.

Somatic Marker Hypothesis

It has been suggested that the orbital PFC is responsible for labeling people, objects, or situations with an "affective significance" (Damasio, 1994). This is achieved by associating memories of past, affectively laden events with a representation of the somatic state that the event evoked. Thus, the memory of the past event is labeled with a "somatic marker" that helps individuals make decisions. Rather than having to evaluate all the pros and cons of prior circumstances to decide on the present, one can simply recall the associated somatic marker, providing the individual with the appropriate affective state and, in effect, a "gut-feeling" of which alternative is correct. Patients with damage to this system might be unable to reach a decision, simply because they become incapacitated by contemplating irrelevant information. None of the outcomes of their actions will be obviously more

emotionally preferable than alternative outcomes, and patients might make decisions randomly or impulsively and as such appear to be disinhibited.

A deficient somatic marker system would explain the impairments that patients with damage to the orbitofrontal PFC exhibit on the "gambling task" (Fig. 50.6D). The reward contingencies associated with each deck of cards are designed to be too complex to be ascertained purely by logical reasoning. Control subjects succeed at the task because they associate a somatic marker with each deck of cards, warning the subject as to whether the choice they are about to make is risky. Further support comes from measures of skin conductance responses (SCRs) that reflect autonomic nervous system processing. Control subjects show SCRs prior to the selection of a card, and choices from risky decks are preceded by a larger SCR than choices from less risky decks. Patients with damage to the ventromedial PFC showed no anticipatory SCRs, although their SCRs when actually receiving the reward or punishment are intact.

This hypothesis is similar to one advanced by William James in the nineteenth century. He proposed that emotional feelings depend on feedback from the autonomic nervous system. For instance, running from a bear because we are afraid evokes a particular autonomic state (such as increased heart rate and blood pressure) that makes us feel afraid. Thus, as James stated, we do not run because we are afraid, but we are afraid because we run. A problem with simply using autonomic nervous system signals, however, is that autonomic states produced by different emotions can be identical. Later studies showed that manipulations of the autonomic nervous system, for example, by administering adrenaline, could alter the magnitude but not necessarily the type of emotion experienced in human subjects. These problems are avoided if the orbitofrontal cortex does not need to access the autonomic nervous system directly, but rather can use a conditioned association between the autonomic state and various cortical representations. For example, associations with the somatosensory cortex may be involved in emotional interpretations; patients with damage to the somatosensory cortex have difficulty recognizing facial emotions as do patients with orbital frontal damage.

Multiple Processing Levels

It has been proposed that the PFC is organized by level of processing, much as there are primary and higher-order sensory cortices (Petrides, 1996). The ventrolateral PFC is primary. It receives inputs from sensory cortex and is responsible for their simple maintenance in working memory and for retrieval of information from long-term memory. The dorsolateral PFC then takes this information and performs more executive processes, such as monitoring multiple behaviors and manipulating information. For example, simply rehearsing a phone number is a ventrolateral PFC function, while saying that same number backward is a dorsolateral PFC function. It should be noted that such an arrangement would preclude the possibility of demonstrating a functional double dissociation between these two areas, since the dorsolateral PFC is only able to perform executive processes on information after it had been received by the ventrolateral PFC. Thus, removing the dorsolateral PFC would impair the higher-level processes, while removing the ventrolateral PFC would impair both.

This hypothesis has been supported by evidence from neuroimaging in humans. Subjects were required to perform five spatial working memory tasks while they were scanned using positron emission tomography. This technique enables the visualization of areas of the brain with the increased blood flow that is related to neural activity. Two of the tasks had simple maintenance requirements; they activated the ventrolateral PFC. The three other tasks also had manipulation requirements and they activated the dorsolateral PFC.

Rule Representation and Hierarchical Organization

Several investigators have pointed out that because knowledge about goals, means, and consequences is vital to planning and implementing complex goal-directed behavior, the PFC is likely to play a central role in its acquisition and representation. The neural instantiation of task demands in the PFC was first proposed by Joaquin Fuster (1989). Based on behavioral and neurophysiological experiments in monkeys, he hypothesized that sustained activity in the PFC was not used merely to retain recent sensory events but also to link together behavioral relevant, but temporally separate, events and actions into a representation of task contingencies.

Other investigators have reached similar conclusions. Grafman (1994) suggested that the PFC contains managerial knowledge units (MKUs), sets of task contingencies linked under common themes that can be flexibly combined into extended routines. An MKU may describe the expected events and appropriate responses for dining in a restaurant, for example. Passingham (1993) suggested that many of the deficits following PFC damage can be explained by a loss of the ability to acquire conditional if-then rules. Wise, Murray, and Gerfen (1996) proposed complementary roles in rule learning for the PFC and the basal ganglia. The PFC is thought to act when new rules are learned while the basal ganglia is thought to potentiate (select) the previously learned rules that are appropriate for the current situation.

Cognitive control has also been studied by developing simulated neural network models that can perform tasks that depend on controlled processing, such as the

Stroop Test. These models invariably include units dedicated to representing task rules and contingencies. Analogies are often drawn between these units and the PFC.

How might the PFC acquire rule information and use it for cognitive control? Miller and Cohen (2001) proposed that reward signals from midbrain dopamine systems act on the multimodal circuitry in the PFC to strengthen connections between neurons that process the information that led to the reward. This action results in a model of the task contingencies, which in neural terms amounts to a representation of the pattern of connections among the sensory, motor, and mnemonic representations needed to solve that task. Maintenance mechanisms in the PFC sustain this representation during task performance. This produces feedback signals to posterior cortex that bias the flow of activity along lines that match the PFC representation, i.e. along task-relevant neural pathways. In short, the PFC plots a "map" of which neural pathways in the forebrain are successful at solving a task and then holds it online in the PFC so that the rest of the neocortex can follow it.

Evidence for this proposal comes from the finding that PFC neurons show neural activity that reflects behaviorally relevant associations between disparate and arbitrary events and actions, behavioral context, and rules. Behavioral and neurophysiological studies of focal attention also point to the operation of top-down signals on the visual cortex (presumed to come from the PFC) that enhance representations of behaviorally relevant stimuli and visual field locations at the expense of irrelevant ones. Evidence that such signals come from the PFC was supported by a recent observation that disconnecting the PFC from the temporal cortex prevented the activation of stored memories in the latter and caused monkeys to fail at a recall test (Tomita, Ohbayashi, Nakahara, Hasegawa, & Miyashita, 1999).

Recent work has begun to look at how the representation of rules is organized in PFC. The picture that is emerging from these studies is that more abstract information is represented by progressively more abstract regions of PFC (Badre & D'Esposito, 2007). Neuroimaging studies show that simple associations between specific stimuli and specific motor responses (e.g., if red light, then stop) are encoded by areas immediately posterior to PFC in the premotor cortex. Lateral PFC regions are activated as rules become more abstract, for example, dependent on abstract stimulus properties such as shape or color. Finally, the most anterior parts of PFC are activated when rules are dependent on the context in which they occur, whether that context is defined temporally, spatially, or relationally.

These findings suggest that rules may be organized hierarchically within PFC. In a hierarchical organization, information at higher levels of the hierarchy is dependent on lower levels. Studies of patients with damage to different parts of PFC suggest that this is, indeed, the case (Badre, Hoffman, Cooney, & D'Esposito, 2009). Damage to anterior PFC only impairs the highest level of the hierarchy (contextual rules). Damage to lateral PFC impairs contextual rules and rules based on abstract stimulus properties, but leaves intact specific-response associations. Finally, damage to premotor cortex impairs all types of rule implementation. Recent computational models have shown how reinforcement learning rules, thought to be encoded by dopamine neurons, could lead to the development of hierarchical representations (Botvinick, Niv, & Barto, 2009; Frank & Badre, 2011).

The Flexibility and Limited Capacity of Controlled Processing

Sustained activation is central in many views of PFC functions. This is not surprising considering the importance of working memory for cognitive control and that sustained activity is common in the PFC. But as O'Reilly and Munakata (2000) have pointed out, representing information through sustained activity offers more than just the ability to bridge gaps; it has other advantages for cognitive control.

It is universally assumed that long-term storage in the brain depends on changing synaptic weights, by strengthening some neural connections and weakening others. Encoding information in structural changes has obvious advantages for long-term storage but the resulting memories are relatively inflexible. Also, changing synaptic weights has local effects; only the two neurons sharing the synapse are affected. The information contained in a set of synaptic weights is thus expressed only when the circuit is fired. By contrast, the information needed to guide goal-directed behavior must be expressed in a format that allows it to affect the ongoing processing in other brain systems. Sustained activity is such a format. It is extended over time and information contained in a pattern of activity can be propagated across the brain. Thus, the ability of sustained activity to tonically influence other brain systems is likely important for coordinating diverse processing around a specific goal. It also affords flexibility; if cognitive control stems from a pattern of information maintained in the PFC, changing behavior is as easy as changing the pattern (O'Reilly & Munakata, 2000; Miller & Cohen, 2001).

Finally, the neuronal mechanisms by which information needed for cognitive control is expressed might explain the severely limited capacity of controlled processes. A vast amount of information can be stored by changing synaptic weights of a network that contains many neurons, each with many potential connections. At any given time, however, most of this information is latent; it is expressed only when the relevant circuit is activated. By contrast, if the information for cognitive control

is expressed in a unique pattern of ongoing activity distributed across many simultaneously active neurons—a population code—then there will be a natural capacity limitation. Trying to represent more than just a few items at the same time would degrade information because the unique patterns impinging on a given set of neurons would overwrite and interfere with one another.

The PFC versus the Basal Ganglia: Model-Building versus "Snapshots"

The prefrontal cortex (and other frontal cortical areas) forms anatomical loops with the basal ganglia (BG), a set of subcortical structures that have motor and learning functions (see Chapter 30). Different regions of the frontal lobe (e.g., the lateral PFC, orbital PFC, premotor cortex) have anatomically segregated loops with the BG that are "closed," meaning that the same region of the frontal cortex that gives rise to the loop is also the target of that loop via BG's connections with the thalamus. Closed loops are noteworthy because they may allow recursive processing in which the results of processing are fed back into the loop for further elaboration. It is possible, if not likely, that some form of recursive processing is responsible for the open-ended nature of advanced thought.

Daw and colleagues (2005) suggested functional specializations for the PFC and BG (specifically the striatum, the BG structure that receives input from the cerebral cortex). They suggested that the PFC builds models of an entire behavior; it retains information about the overall structure of the task, following the whole course of action from initial state to ultimate outcome. They liken this to a "tree" structure for a typical operant task: behaviors begin in an initial state with two or more possible response alternatives. Choosing one response leads to another state with new response alternatives, with this process continuing throughout the task, ultimately leading to a reward. The PFC is able to capture this entire tree structure, essentially providing the animal with an internal model of the task. By contrast, the striatum is thought to learn the task piecemeal, with each state's response alternatives individually captured as separate "frames." This "caching reinforcement learning" system retains information about which alternative is "better" in each state but nothing about the overall structure of the task (i.e., the whole tree). In short, the BG learns each fork in the road, while the PFC puts together the whole route.

This is thought to explain observations of tasks that use reinforcer devaluation. In such tasks, you change the value of the reward by saturating the animal on a given reward (for example, overfeeding on chocolate if chocolate is a reward on that task). This has revealed two classes of behavior. Behaviors that are affected by reinforcer devaluation are considered to be goal-directed because changing the goal changes the behavior. As discussed above, goal-directed behaviors depend on the PFC. By contrast, overlearned behaviors whose outcomes remain relatively constant can become habits, impervious to reinforcer devaluation. Because these behaviors are not affected by changing the goal, they seem to reflect control by a frame-based caching system in which the propensity for a given alternative in each situation is stored independently of information about past or future states. Indeed, such a piecemeal caching system could only acquire complex tasks as habits because it can only learn them if they are always conducted in the same fashion. Habits have long been thought to be a specialization of the BG.

Support for this model comes from recent observations that during conditional visuomotor rule learning, the dorsal striatum of the BG (the portion that receives direct projections from the lateral PFC and whose output ultimately loops back to the lateral PFC) showed very fast, almost bi-stable changes in neural activity (Pasupathy & Miller, 2005). By contrast, the lateral PFC showed slower gradual changes in neural activity. This may reflect the much heavier input from midbrain dopaminergic reward-related neurons in the striatum than the PFC. The fast plasticity in the striatum could be ideal for learning the reinforcement-related snapshots that capture the immediate circumstances and which alternative is preferable for a particular state. The slow plasticity in the PFC is better suited for the linking in of additional information about past states that is needed to learn and retain an entire model of the task and thus predict future states.

Top-Down Control of Posterior Cortex by the PFC

The theories discussed above all either tacitly or explicitly assume that the PFC is the provider of top-down control that modulates processing in other brain areas biasing it toward task-relevant information. Depending on the brain system that receives those signals, the result could be attention, memory retrieval, or motor control. Due to the technological difficulty in directly measuring how different brain structures interact, most of the evidence for this has been indirect. Recent work, however, has provided some direct evidence, mainly from studies of visual attention.

For example, Tirin Moore and colleagues have found that stimulating FEF neurons at subthreshold levels (i.e., levels that would not elicit an eye movement) induces attention-like effects in V4 neurons whose receptive fields overlap with the FEF stimulation sites (Moore & Armstrong, 2003). Dopamine may play an important role in the allocation of attention, as stimulating dopamine D1R receptors in FEF leads to attention-like effects (Noudoost & Moore, 2011). These results suggest that

FEF, which is known to play a role in generating volitional eye movements, may also play a direct role in top-down direction of attention.

Further support comes from simultaneous multielectrode recording from the PFC and posterior cortex. Multiple electrodes allow detailed examination of neural latencies and synchrony between neurons that gives insight into network properties and the flow of neural activity (Miller & Wilson, 2008). For example, both the PFC and an area of the parietal cortex called the lateral intraparietal area (LIP) are thought to play roles in focusing visual attention. Simultaneous recording from them while monkeys volitional searched a display for a target object (i.e., in a top-down fashion), neurons in the PFC reflected the shift of attention to the target before LIP (Buschman & Miller, 2007). By contrast, when the target was more salient than other stimuli and the monkey's attention was reflexively drawn to it (i.e., in a bottom-up or sensory-driven fashion), the shift of attention to the target was reflected first in LIP. The timing of appearance of these signals in the two areas suggests that the top-down control signals originated in the prefrontal cortex and then flowed posteriorly to LIP whereas bottom-up sensory signals flowed the opposite direction (Fig. 50.10).

Synchronized, rhythmic oscillations between brain areas may also play a role in top-down control. These "brain waves" are thought, among other things, to facilitate long-range communication between brain areas. When neurons are simultaneously depolarized, they are more susceptible to influence from each other (conversely, when they are out of phase, they are less likely to communicate). Long-range oscillatory coherence is found, for example, between the PFC and LIP during shifts of attention (Buschman & Miller, 2007). Desimone and colleagues paired recordings at the posterior end of the PFC, in the frontal eye fields (FEF), with visual cortical area V4 and found that focusing attention on a visual stimulus increased high frequency coupling between them (Gregoriou, Gotts, Zhou, & Desimone, 2009). The time-shift of their coupling suggested that the coupling was initiated by the FEF. Such long-range oscillations may dynamically sculpt subnetworks out of the greater number of anatomical connections. Thus, they may help direct top-down signals to specific neuron populations.

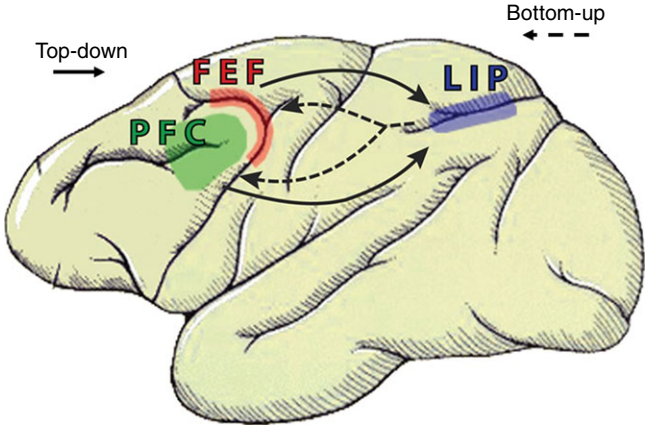


FIGURE 50.10 Schematic representation of the flow of top-down and bottom-up attention signals in the PFC. *Based on neural latency from Buschman and Miller (2007).*

Summary

The ability to take charge of one's actions and direct them toward future aims is called cognitive control. Virtually all theories of cognition posit that cognition depends on functions specialized for the acquisition of information about goals and means. These functions exert a top-down influence on the lower-level automatic processes that mediate sensory analysis, memory storage, and motor outputs, orchestrating and directing them toward a given goal.

The PFC, a brain structure that reaches its greatest complexity in the primate brain, seems to play a central role in cognitive control. It has the requisite anatomical infrastructure. It has access to, and the means to influence processing in, all major forebrain systems and can provide a means to synthesize the diverse information related to a given goal. Indeed, its damage in humans and monkeys results in a dysexecutive syndrome; individuals are stimulus-bound and seem capable only of reacting to the immediate environment. Information about unseen goals and potential consequences of actions has little influence on behavior. PFC neurons also have properties consistent with their participation in cognitive control: they are multimodal, reflect task demands, and can sustain their activity to keep task-relevant information online during task performance.

Theories of PFC function suggest that it makes several contributions important for cognitive control: maintenance and manipulation of information in working memory, the assignment and recall of the affective tags related to different choices, and the acquisition and representation of behavior-guiding rules that help orchestrate goal-directed behaviors. None of these theories is mutually exclusive, and indeed, different theories may be focusing on different aspects of a complex system that is involved in a wide range of brain processes.

References

Asaad, W. F., Rainer, G., & Miller, E. K. (1998). Neural activity in the primate prefrontal cortex during associative learning. *Neuron, 21,* 1399–1407.

Asaad, W. F., Rainer, G., & Miller, E. K. (2000). Task-specific activity in the primate prefrontal cortex. *Journal of Neurophysiology, 84,* 451–459.

Badre, D., & D'Esposito, M. (2007). Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *Journal of Cognitive Neuroscience, 19*(12), 2082–2099.

Badre, D., & D'Esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature Reviews Neuroscience, 10*(9), 659–669.

Badre, D., Hoffman, J., Cooney, J. W., & D'Esposito, M. (2009). Hierarchical cognitive control deficits following damage to the human frontal lobe. *Nature Neuroscience, 12*(4), 515–522.

Bichot, N. P., & Schall, J. D. (1999). Effects of similarity and history on neural mechanisms of visual selection. *Nature Neuroscience, 2,* 549–554.

Botvinick, M. M., Niv, Y., & Barto, A. C. (2009). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition, 113*(3), 262–280.

Buschman, T. J., & Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science, 315,* 1860–1862.

Cromer, J. A., Roy, J. E., & Miller, E. K. (2010). Representation of multiple, independent categories in the primate prefrontal cortex. *Neuron, 66,* 796–807.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York, NY: Putman.

Damasio, H., Grabowski, T., Frank, R., Galaburda, A. M., & Damasio, A. R. (1994). The return of Phineas Gage: clues about the brain from the skull of a famous patient. *Science, 264,* 1102–1105.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci, 8,* 1704–1711.

Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive Psychology, 30,* 257–303.

Fellows, L. K., & Farah, M. J. (2007). The role of ventromedial prefrontal cortex in decision making: Judgment under uncertainty or judgment per se? *Cerebral Cortex, 17*(11), 2669–2674.

Fitzgerald, J. K., Freedman, D. J., & Assad, J. A. (2011). Generalized associative representations in parietal cortex. *Nature neuroscience 14,* 1075–1079.

Frank, M. J., & Badre, D. (2011). Mechanisms of hierarchical reinforcement learning in corticostriatal circuits 1: Computational analysis. *Cerebral Cortex*.

Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology, 61,* 331–349.

Fuster, J. M. (1973). Unit activity in prefrontal cortex during delayed-response performance: neuronal correlates of transient memory. *Journal of Neurophysiology, 36,* 61–78.

Fuster, J. M. (1989). *The prefrontal cortex*. New York, NY: Raven Press.

Goldman-Rakic, P. S. (1987). Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In F. Plum (Ed.), *Handbook of physiology: The nervous system* (pp. 373–417). Bethesda, MD: American Physiological Society.

Grafman, J. (1994). Alternative frameworks for the conceptualization of prefrontal functions. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (pp. 187). Amsterdam: Elsevier.

Gregoriou, G. G., Gotts, S. J., Zhou, H., & Desimone, R. (2009). High-frequency, long–range coupling between prefrontal and visual cortex during attention. *Science, 324,* 1207–1210.

Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science, 302*(5648), 1181–1185.

Kubota, K., & Niki, H. (1971). Prefrontal cortical unit activity and delayed alternation performance in monkeys. *Journal of Neurophysiology, 34,* 337–347.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal function. *Annual Review of Neuroscience, 24,* 167–202.

Miller, E. K., & Wilson, M. A. (2008). All my circuits: Using multiple-electrodes to understand functioning neural networks. *Neuron, 60,* 483–488.

Moore, T., & Armstrong, K. M. (2003). Selective gating of visual signals by microstimulation of frontal cortex. *Nature, 421,* 370–373.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research and theory* (pp. 1–48). New York, NY: Plenum.

Noudoost, B., & Moore, T. (2011). Control of visual cortical signals by prefrontal dopamine. *Nature, 474,* 372–375.

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind*. Cambridge: MIT Press.

Padoa-Schioppa, C., & Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature, 441,* 223–226.

Passingham, R. (1993). *The frontal lobes and voluntary action*. Oxford: Oxford Univ. Press.

Pasupathy, A., & Miller, E. K. (2005). Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature, 433,* 873–876.

Petrides, M. (1996). Specialized systems for the processing of mnemonic information within the primate frontal cortex. *Philosophical Transactions of the Royal Society of London. B:Biological Sciences, 351,* 1455–1461.

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience, 9*(7), 545–556.

Rigotti, M., Rubin, D. B., Wang, X. J., & Fusi, S. (2010). Internal representation of task rules by recurrent dynamics: the importance of the diversity of neural responses. *Frontiers in Computational Neuroscience, 4,* 24.

Tomita, H., Ohbayashi, M., Nakahara, K., Hasegawa, I., & Miyashita, Y. (1999). Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature, 401,* 699–703.

Wallis, J. D. (2007). Neuronal mechanisms in prefrontal cortex underlying adaptive choice behavior. *Annals of the New York Academy of Sciences, 1121,* 447–460.

Wallis, J. D., & Kennerley, S. W. (2010). Heterogeneous reward signals in prefrontal cortex. *Current Opinion in Neurobiology, 20*(2), 191–198.

Wallis, J. D., Anderson, K. C., & Miller, E. K. (2001). Single neurons in the prefrontal cortex encode abstract rules. *Nature, 411,* 953–956.

Walton, M. E., Behrens, T. E., Buckley, M. J., Rudebeck, P. H., & Rushworth, M. F. (2010). Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. *Neuron, 65*(6), 927–939.

Wise, S. P. (2008). Forward frontal fields: Phylogeny and fundamental function. *Trends in Neurosciences, 31*(12), 599–608.

Wise, S. P., Murray, E. A., & Gerfen, C. R. (1996). The frontal-basal ganglia system in primates. *Critical Reviews in Neurobiology, 10,* 317–356.

Suggested Readings

Adolphs, R. (2001). The neurobiology of social cognition. *Current Opinion in Neurobiology, 11,* 231–239.

Fuster, J. M. (2001). The prefrontal cortex–an update: Time is of the essence. *Neuron, 30,* 319–333.

Earl K. Miller and Jonathan D. Wallis

This page intentionally left blank

CHAPTER

51

The Neuroscience of Consciousness

Consciousness is one of the most enigmatic features of the universe. People—and presumably many other animals, too—not only act but feel: they see, hear, smell, recall, plan for the future. These activities are associated with subjective, ineffable, immaterial feelings that are tied in some manner to the material brain. The exact nature of this relationship—the classical mind-body problem—remains elusive and the subject of heated debate. These firsthand, subjective experiences pose a daunting challenge to the scientific method that has, in many other areas, proven so immensely fruitful. Science can describe events microseconds following the Big Bang, offer an increasingly detailed account of matter and how to manipulate it, and uncover the biophysical and neurophysiological nuts and bolts of the brain and its pathologies. However, this same method has as yet failed to provide a satisfactory account of how firsthand, subjective experience fits into the objective, physical universe. The brute fact of consciousness comes as a total surprise; it does not appear to follow from any phenomena in physics or biology. Indeed, some modern philosophers argue that consciousness is not logically supervenient to physics (Chalmers, 1996). That is, it is certainly logically possible to imagine a world identical to ours but without any conscious experience. Nothing in the known laws of physics—or of biology—seems to contradict this. Yet we do find ourselves in a world where each one of us has conscious experiences, sees a picture of the world and so on.

People willingly concede that when it comes to nuclear physics or molecular biology, specialist knowledge is essential; but many assume that there are few relevant facts about consciousness and therefore everybody is entitled to their own theory. Nothing could be further from the truth. There is an immense amount of relevant psychological, clinical, and neuroscientific data and observations that must be accounted for. Furthermore, the modern focus on the neuronal basis of consciousness in the brain—rather than on interminable philosophical debates—has given brain scientists tools to greatly increase our knowledge of the conscious mind.

Consciousness is a state-dependent property of certain types of complex, biological, adaptive, and highly interconnected systems. The best example of consciousness is found in a healthy and attentive human brain—for example, the reader of this chapter. Not all biological, adaptive systems appear to have consciousness; examples include the enteric nervous system and the immune system. Brain scientists are exploiting a number of empirical approaches that shed light on the neural basis of consciousness. This chapter reviews these approaches and summarizes what has been learned.

WHAT PHENOMENA DOES CONSCIOUSNESS ENCOMPASS?

There are many definitions of *consciousness* (Searle, 2004). A common philosophical one is "*Consciousness is what it is like to be something*," such as the experience of what it feels like to see red, to be me, or to be angry. This what-it-feels-like-from-within definition expresses the principal irreducible characteristic of the phenomenal aspect of consciousness: *to experience something*. "What it feels like to see red" also emphasizes the subjective or *first-person perspective* of consciousness: it is a subject, an I, who is having the experiences, and the experience is inevitably private.

A science of consciousness must explain the exact relationship between phenomenal, mental states, and brain states. This is the heart of the classical mind-body

© 2013 Elsevier Inc. All rights reserved.

problem: *What is the nature of the relationship between the immaterial, conscious mind and its physical basis in the electrochemical interactions in the body?* This problem can be divided into several subproblems:

1. Why is there any experience at all? Or, put differently, why does any one brain state—that is, neurons over here firing, while neurons over there remain silent—feel like anything? In philosophy, this is referred to by some as the *Hard Problem* (note the capitalization), or as the explanatory gap between the material, objective world and the subjective, phenomenal world (Chalmers, 1996). Many scholars have argued that the exact nature of this relationship will remain a central puzzle of human existence, without an adequate reductionistic, scientific explanation. However, as similar sentiments have been expressed in the past for the problem of seeking to understand life or to determine what material the stars are made out of, it is best to put this question aside for the moment and not be taken in by defeatist arguments.
2. Why is the relationship among different experiences the way it is? For instance, red, yellow, green, cyan, blue, and magenta are all colors that can be mapped onto the topology of a circle. Why? Furthermore, as a group, these color percepts share certain commonalities that make them different from other percepts, such as seeing motion or smelling a rose.
3. Why are feelings private? As expressed by poets and novelists, we cannot communicate an experience to somebody else except by way of example. Try explaining vision to somebody blind from birth, or color to an achromat—rare individuals born without any cone photoreceptors.
4. How do feelings acquire meaning? Subjective states are not abstract states but have an immense amount of associated explicit and implicit feelings. Think of the unmistakable smell of dogs coming in from the rain or the crunchy texture of potato chips.
5. Why are only some behaviors associated with conscious states? Much brain activity and associated behavior occur without any conscious sensation.
6. What is the function—if any—of conscious experiences?

THE NEUROBIOLOGY OF FREE WILL

A further aspect of the mind-body problem is the question of free will, a vast topic. Answering this question goes to the heart of the way people think of themselves. The spectrum of views ranges from the traditional and deeply embedded belief that we are free, autonomous, and conscious actors to the view that we are biological machines driven by needs and desires beyond conscious access and with limited insight or voluntary control.

Of great relevance are the classical findings by Libet, Gleason, Wright, and Pearl (1983) of brain events that precede the conscious initiation of a voluntary action. In this seminal experiment, subjects were sitting in front of an oscilloscope, tracking a spot of light moving every 2.56 seconds around a circle. Every now and then, "spontaneously," the subject had to carry out a specific voluntary action, here flexing their wrist. If this action is repeated sufficiently often while electrical activity around the vertex of the head is recorded, a *readiness potential* (*Bereitschaftspotential*) in the form of a sustained scalp negativity develops 1–2 sec before the muscle starts to move. Libet asked subjects to silently note the position of the spot of light when they first "felt the urge" to flex their wrist and to report this location afterward. This temporal marker for the awareness of willing an action occurs on average 200 msec before initiation of muscular action (with a standard error of about 20 msec), in accordance with commonsense notions of the causal action of free will. However, the readiness potential can be detected at least 350 msec before awareness of the action. In other words, the subject's brain signals the action at least half a second before the subject feels that he or she has initiated it!

The original experiment has been replicated (Haggard & Eimer, 1999) as well as extended using fMRI, where precentral and parietal cortices can signal decisions up to ten seconds prior to entering awareness. Yet, because of its counterintuitive implication that conscious will has no causal role, the interpretation of these experiments continues to be vigorously debated (Haggard, 2008).

Psychological work in normal individuals and in patients reveals further dissociations between the conscious perception of a willed action and its actual execution: subjects believe that they perform actions that they did not do, while, under different circumstances, subjects feel that they are not responsible for actions that are, demonstrably, their own (Haggard, 2008; Wegner, 2002).

Yet whether volition is illusory or is free in some libertarian sense does not answer the question of how subjective states relate to brain states. The perception of free will, what psychologists call the *feeling of agency* or *authorship* (e.g., "I decided to lift my finger"), is a subjective state with an associated quale no different in kind from the quale of a toothache or seeing marine blue. So even if free will is a complete chimera, the subjective feeling of willing an action must have some neuronal correlate.

The quiddity of the sensation of agency has been strengthened by direct electrical brain stimulation during neurosurgery and fMRI experiments. These implicate posterior parietal, medial pre-motor and anterior cingulate cortices in generating the subjective feeling of triggering an action (Desmurgert et al., 2009; Haggard, 2008). In other words, the neural correlate for the feeling of apparent causation involves activity in these regions.

CONSCIOUSNESS IN OTHER SPECIES

Data about subjective states come not only from people who can talk about their subjective experiences but also from nonlinguistic competent individuals—such as newborn babies or patients with complete paralysis of nearly all voluntary muscles (locked-in syndrome)—and, most importantly, from animals other than humans. There are three reasons to assume that many species, in particular those with complex behaviors such as mammals, share at least some aspects of consciousness with humans:

1. *Similar neuronal architecture:* Except for size, there are no large-scale genomic, architectonic, cellular, or neurophysiological differences between the cerebral cortex and thalamus of mice, monkeys, humans, and whales. They are not the same and obey slightly different anatomical scaling laws but are overall remarkably similar (Herculano-Houzel, 2009).
2. *Similar behavior:* Almost all human behaviors have precursors in the animal literature. Take the case of pain. The behaviors seen in humans when they experience pain and distress—facial contortions, moaning, yelping, or other forms of vocalization; motor activity such as writhing; avoidance behaviors at the prospect of a repetition of the painful stimulus—can be observed in all mammals and in many other species. Likewise for the physiological signals that attend pain—activation of the sympathetic autonomous nervous system, resulting in change in blood pressure, dilated pupils, sweating, increased heart rate, release of stress hormones, and so on. The discovery of cortical pain responses in premature babies (Slater et al., 2006) and of empathy and observational pain in mice (Jeon et al., 2010) shows the fallacy of relying on language as sole criteria for consciousness.
3. *Evolutionary continuity:* The first true mammals appeared at the end of the Triassic period, about 220 million years ago, with primates proliferating following the Cretaceous-Tertiary extinction event, about 60 million years ago, while humans and macaque monkeys did not diverge until 30 million years ago (Allman, 1999). *Homo sapiens* is part of an evolutionary continuum with its implied structural and behavioral continuity, rather than an independently developed organism.

While certain aspects of consciousness, in particular those relating to the self, introspection and to abstract, culturally transmitted knowledge, may not be widespread in nonhuman animals, there is little reason to doubt that other mammals share conscious feelings—sentience—with humans. To believe that humans are special, are singled out by the gift of consciousness above all other species, is a remnant of humanity's atavistic, deeply held belief that *Homo sapiens* occupies a privileged place in the universe, a belief with no empirical basis.

The extent to which nonmammalian vertebrates, such as tuna, cichlid, and other fish; crows, ravens, magpies, parrots, and other birds; or even invertebrates, such as squids or bees, with complex, nonstereotyped behaviors including delayed-matching, non-matching-to-sample, and other forms of learning (Giurfa, Zhang, Jenett, Menzel, & Srinivasan, 2001) are conscious is difficult to answer at this point in time. Without a sounder understanding of the neuronal architecture necessary to support consciousness, it is unclear where in the animal kingdom to draw the Rubicon that separates species with at least some conscious percepts from those that never experience anything and that are nothing but pure automata (Edelman, Baars, & Seth, 2005, Koch, 2012).

AROUSAL AND STATES OF CONSCIOUSNESS

There are two quite distinct meanings of *being conscious*, relating to the intransitive and to the transitive usage of the verb. The former revolves around *arousal* and *states of consciousness*, while the latter deals with *the content of consciousness* and *conscious states*.

To be conscious of anything, the brain must be in a relatively high state of arousal. This is as true of wakefulness as it is of REM sleep that can be vividly, consciously experienced—though usually not remembered—in dreams (see Chapter 40). The level of brain arousal, measured by electrical or metabolic brain activity, fluctuates in a circadian manner, and is influenced by lack of sleep, drugs and alcohol, physical exertion, and so on in a predictable manner. High arousal states are always associated with some conscious state—a percept, thought, or memory—that has a specific *content.* We see and hear something, remember an incident, plan the future, or fantasize about sex. Indeed, it is not clear whether one can be awake without being conscious of something (with the possible exception of certain meditative states). Referring to such conscious states is conceptually quite distinct from referring to states of consciousness that fluctuate with different levels of arousal. Arousal can be measured behaviorally by the signal amplitude that triggers some criterion reaction (for instance, the sound level necessary to evoke an eye movement or a head turn toward the sound source).

The number of distinct conscious states is vast and encompasses every possible type of visual, auditory, visceral, olfactory, pain, vestibular, or other sensory experience; remembered or imagined experiences; events or thoughts; and emotional experiences (fear, surprise, happiness, dread, and so on).

Different levels or states of consciousness are associated with different kinds of conscious experiences. The normal, waking state is quite different from the dreaming state (for instance, the latter has little or no self-reflection) or from deep sleep. In all three cases, the basic physiology of the brain is changed, affecting the space of possible conscious experiences. Physiology is also different in *altered states of consciousness*—for instance, after taking psychedelic drugs, when events often have a stronger emotional connotation than in normal life.

In some obvious but difficult to rigorously define manner, the *richness of conscious experience* increases as an individual transitions from deep sleep to drowsiness to full wakefulness. This richness of possible conscious experience could be quantified using notions from complexity theory that incorporate both the dimensionality as well as the granularity of conscious experience (e.g., Tononi, 2008). For example, inactivating all of the visual cortex in an otherwise normal individual would significantly reduce the dimensionality of conscious experience, since no color, shape, motion, texture, or depth could be perceived or imagined. As behavioral arousal increases, so does the range and complexity of behaviors that an individual is capable of. A singular exception to this progression is REM sleep where most motor activity is shut down in the *atonia* that is characteristic of this phase of sleep, and the subject is difficult to wake up. Yet this low level of behavioral arousal goes, paradoxically, hand in hand with high metabolic and electrical brain activity and conscious, vivid states.

These observations suggest a two-dimensional graph (Fig. 51.1) in which the richness of conscious experience (its representational capacity) is plotted as a function of levels of behavioral arousal or responsiveness.

Global disorders of consciousness can likewise be mapped onto this plane. Clinicians speak of impaired states of consciousness as in "the comatose state," "the vegetative state" (VS), and the "minimal conscious state" (MCS). Here, *state* refers to different levels of consciousness, from a total absence in the case of coma, VS, or general anesthesia, to a fluctuating and limited form of conscious sensation in MCS, sleepwalking, or during a complex partial epileptic seizure (Laureys, 2005; Schiff, 2010).

The repertoire of distinct conscious states or experiences that are accessible to a patient in MCS is presumably minimal (possibly including pain, discomfort, and sporadic sensory percepts), immeasurably smaller than the possible conscious states that can be experienced by a healthy brain. In the limit of brain death, the origin of this space has been reached with no experience at all (Fig. 51.1). More relevant to clinical practice is the case of global anesthesia, during which the patient should not experience anything so as to avoid traumatic memories and their undesirable sequelae.

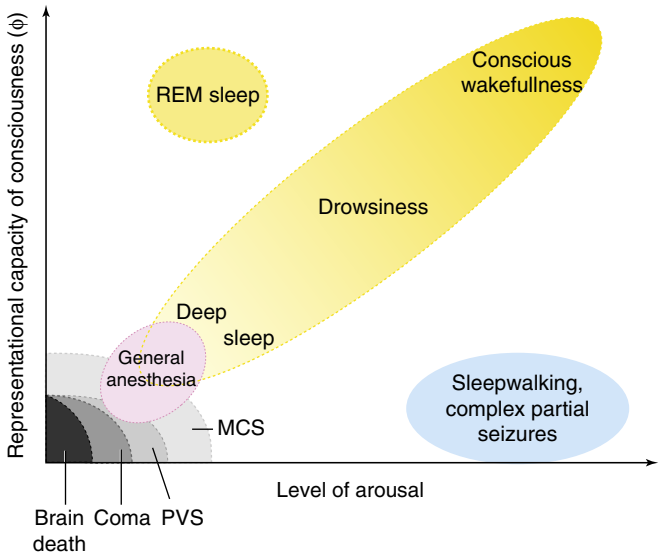


FIGURE 51.1 Normal and pathological brain states can be situated in a two-dimensional graph. Increasing levels of behaviorally determined arousal are plotted on the *x*-axis and the "richness" or "representational capacity of consciousness" is plotted on the *y*-axis. Increasing arousal can be measured by the threshold to obtain some specific behavior (for instance, spatial orientation to a sound). Healthy subjects cycle during a 24-hour period from deep sleep with low arousal and very little conscious experience to increasing levels of arousal and conscious sensation. In REM sleep, low levels of behavioral arousal go hand-in-hand with vivid consciousness. Conversely, various pathologies of clinical relevance are associated with little to no conscious content. *Modified from Laureys (2005).*

Given the absence of any accepted theory for the minimal neuronal criteria necessary for consciousness, the distinction between a VS patient—who shows regular sleep-wave transitions and who may be able to move eyes or limbs or smile in a reflexive manner as in the widely publicized 2005 case of Terri Schiavo in Florida—and a MCS patient who can communicate (on occasion) in a meaningful manner (for instance, by differential eye movements) and who shows some signs of consciousness, can be difficult in a clinical setting. This drives the need for practical, bedside tests. Two empirico-computational assays being developed rely on the breakdown of cortico-cortical interconnectivity—that is, the loss of integration within the cortico-thalamic core, associated with loss of consciousness (Boly et al., 2011; Massimini et al., 2005; Rosanova et al., 2012).

An alternative is blood-oxygen-level-dependent (BOLD) functional magnetic resonance imaging (fMRI). Brain imaging of patients with global disturbances of consciousness (including akinetic mutism) reveals that dysfunction in a widespread cortical network including medial and lateral prefrontal cortex and parietal associative areas is associated with a global loss of consciousness (Laureys, 2005). Monti et al. (2010) imaged brain activity in 54 patients

with disorders of consciousness to determine whether two different mental imagery tasks that involve different brain systems ("imagine playing tennis" versus "imagine navigating through your home") can be used by the patient to communicate yes or no answers to simple questions. Five of these patients with traumatic brain injuries could willfully and reliably modulate their brain activity, with two of these having no overt sign of behavioral control during clinical assessment. This shows the promise of fMRI for establishing a two-way radio link (albeit at a very low baud rate) with a subset of VS patients.

Destruction of circumscribed parts of the cerebral cortex can eliminate very specific aspects of consciousness, such as the ability to see moving stimuli or to recognize faces, without a concomitant loss of consciousness or even conscious vision in general. Yet relatively discrete bilateral injuries to midline (paramedian) subcortical structures can cause a complete loss of consciousness. These structures are therefore part of the *enabling* factors that control the level of brain arousal (as determined by metabolic or electrical activity) and that are needed for any conscious states to form. Consider the heterogeneous collection of more than 20 (on each side) nuclei in the upper brainstem (pons, midbrain, and in the posterior hypothalamus), often collectively referred to as the *reticular activating system* (Parvizi and Damasio, 2001). These nuclei—three-dimensional collections of neurons with their own cytoarchitecture and neurochemical identity—release distinct neuromodulators such as acetylcholine, noradrenaline/norepinephrine, serotonin, histamine, and orexin/hypocretin. They can widely influence the cortex through direct axonal projections (Fig. 51.2) or via intermediary relays in the thalamus (below) and in the basal forebrain. These neuromodulators mediate the alternation between wakefulness and sleep, as well as the general level of both behavioral and brain arousal. Acute lesions in the reticular activating system can result in loss of consciousness and coma. However, eventually the excitability of thalamus and forebrain can recover and consciousness can return (Villablanca, 2004).

Another enabling factor for consciousness is the five or more *intralaminar nuclei of the thalamus* (ILN). These receive input from many brainstem nuclei and from frontal cortex and project strongly to the basal ganglia and, in a more distributed manner, into layer I of much of neocortex. Comparatively small (1 cm^3 or less) bilateral lesions in the ILN can completely eliminate awareness (Bogen, 1995). Thus, the ILN are necessary for the state of consciousness but do not appear to be responsible for mediating specific conscious percepts or memories.

If a single substance is critical for consciousness, then acetylcholine is the most likely candidate. Two

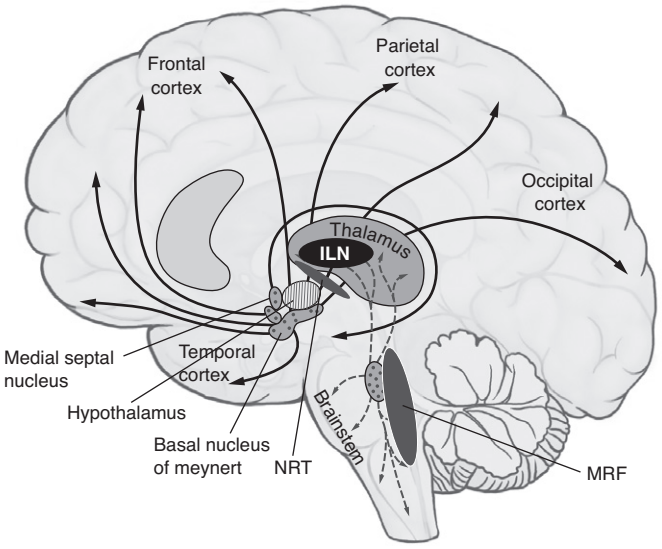


FIGURE 51.2 Midline structures in the brainstem and thalamus necessary to regulate the level of brain arousal include the intralaminar nuclei of the thalamus (ILN), the thalamic reticular nucleus (NRT) encapsulating the dorsal thalamus, and the midbrain reticular formation (MRF) that includes the reticular activating system. Small, bilateral lesions in many of these nuclei cause a global loss of consciousness.

major cholinergic pathways originate in the brainstem and in the basal forebrain (Fig. 51.2). Brainstem cells send an ascending projection to the thalamus, where release of acetylcholine facilitates thalamo-cortical relay cells and suppresses inhibitory interneurons. Cholinergic cells are therefore well positioned to influence all of the cortex by controlling the thalamus. In contrast, cholinergic basal forebrain neurons send their axons to a wide array of target structures. Collectively, brainstem and basal forebrain cholinergic cells innervate the thalamus, hippocampus, amygdala, and neocortex.

Cholinergic activity fluctuates with the sleep-wake cycle. In general, increasing levels of spiking activity in cholinergic neurons are associated with wakefulness or REM sleep, while decreasing levels occur during non-REM or slow-wave sleep. Lastly, many neurological pathologies whose symptoms include disturbances of consciousness, such as Parkinson's disease, Alzheimer's disease, and other forms of dementia, are associated with a selective loss of cholinergic neurons.

In summary, a plethora of nuclei with distinct chemical signatures in the thalamus, midbrain, and pons must function for the brain of an individual to be sufficiently aroused to experience anything at all. These nuclei belong to the enabling factors for consciousness. It is likely that the specific content of any one conscious sensation is mediated by neurons in cortex and their associated satellite structures, including the amygdala, thalamus, claustrum, and the basal ganglia.

THE NEURONAL CORRELATES OF CONSCIOUSNESS

Progress in addressing the mind-body problem has come from focusing on empirically accessible questions rather than on eristic philosophical arguments with no clear resolution. One key objective has been to search for the neuronal correlates—and ultimately the causes—of consciousness. As defined by Crick and Koch (2003), the neuronal correlates of consciousness (NCC) are the *minimal neuronal mechanisms jointly sufficient for any one specific conscious percept* (Fig. 51.3).

This definition of the NCC stresses the word "minimal" because the question of interest is which subcomponents of the brain are actually needed. For example, of the 86 billion neurons in the human brain, 69 billion are in the cerebellum (Herculano-Houzel, 2009). That is, about four out of every five brain cells are small cerebellar granule neurons. Yet the main deficits of people born without a cerebellum—a rare occurrence—or that lose part of the cerebellum to stroke or other trauma are ataxia, slurred speech, and unsteady gait (Lemon & Edgley, 2010). Thus, the cerebellum is not part of the NCC. That is, trains of spikes in granule cells—or their absence—do not contribute in any substantive way to conscious states.

This definition does not focus on the necessary conditions for consciousness because of the great redundancy and parallelism found in neurobiological networks. While activity in some population of neurons may underpin a percept in one case, a different population might mediate a related percept if the former population is lost or inactivated.

Every phenomenal, subjective state will have associated NCC: one for seeing a red patch, another one for seeing grandmother, yet a third one for hearing a siren, and so on. Perturbing or inactivating the NCC for any one specific conscious experience will affect the percept or cause it to disappear. If the NCC could be induced artificially—for instance, by optogenetics or by cortical microstimulation—the subject would experience the associated percept.

What characterizes the NCC? What are the communalities between the NCC for seeing and for hearing? Will the NCC involve all pyramidal neurons in cortex at any given point in time? Or only a subset of long-range pyramidal cells in frontal lobes that project to the sensory cortices in the back? Only layer 5 cortical cells? Neurons that fire in a rhythmic manner? Neurons that fire in a synchronous manner? These are some of the proposals that have been advanced over the years (Chalmers, 2000).

The vast majority of experiments concerned with the NCC refer to either electrophysiological recordings in monkeys or functional brain imaging data in humans.

It is implicitly assumed by most neurobiologists that the relevant variables giving rise to consciousness are to be found at the neuronal level, among the synaptic release or the action potentials in one or more population of cells, rather than at the molecular level. A few scholars have proposed that macroscopic quantum behaviors underlie consciousness. Of particular interest here is *entanglement*, the observation that the quantum states of multiple objects, such as two coupled electrons, may be highly correlated even though they are spatially separated, violating our intuition about locality (entanglement is also the key feature of quantum mechanics exploited in quantum computers). The role of quantum mechanics for the photons received by the eye and for the molecules of life is not controversial. But there is no evidence that any components of the nervous system—a 37° Celsius wet and warm tissue strongly coupled to its environment—display quantum entanglement (although there is evidence for quantum coherence in photosynthetic proteins at room temperature; Collini et al., 2010). And even if quantum entanglement were to occur inside individual cells, diffusion and action

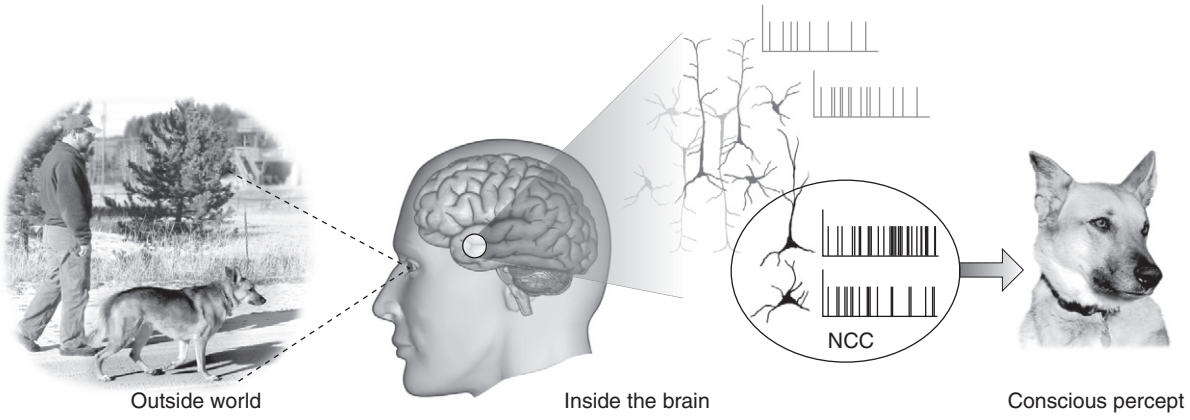


FIGURE 51.3 The **N**euronal **C**orrelates of **C**onsciousness (NCC) are the minimal set of neural events and structures—here synchronized action potentials in neocortical pyramidal neurons—sufficient for a specific conscious percept or memory. *From Koch (2004).*

potential generation and propagation—the principal mechanism for getting information into and out of neurons—would destroy superposition. At the cellular level, the interaction of neurons is governed by classical physics (Koch & Hepp, 2010).

THE NEURONAL BASIS OF PERCEPTUAL ILLUSIONS

The possibility of precisely manipulating visual percepts in time and space has made vision a preferred modality for seeking the NCC. Psychologists have perfected a number of techniques—masking, binocular rivalry, continuous flash suppression, motion-induced blindness, change blindness, inattentional blindness—in which the seemingly simple and unambiguous relationship between a physical stimulus in the world and its associated percept in the privacy of the subject's mind is disrupted (Kim & Blake, 2005). With such techniques, a stimulus can be perceptually suppressed for seconds or even minutes at a time: the image is projected into one of the observer's eyes but it is invisible, not seen. In this manner the neural mechanisms that respond to the subjective percept rather than the retinal stimulus can be isolated, permitting the footprints of visual consciousness to be tracked in the brain.

In a *perceptual illusion*, the physical stimulus remains fixed, while the percept fluctuates. One example is the *Necker cube* that can be perceived in one of two different ways in depth. Another one, that can be precisely controlled, is *binocular rivalry* (Blake & Logothetis, 2002). Here, a small image—for example, a horizontal grating—is presented to the left eye and another image—for example, a vertical grating—is shown to the corresponding location in the right eye. Despite the constant retinal stimulus, observers perceive the horizontal grating to alternate every few seconds with the vertical one. The brain does not allow for the simultaneous perception of both images.

Macaque monkeys can be trained to report whether they see the left or the right image. The distribution of the switching times and the way in which changing the contrast in one eye affects the reports leave little doubt that monkeys and humans experience the same basic phenomenon.

In a series of elegant experiments, Logothetis and colleagues (Logothetis, 1998) recorded from a variety of visual cortical areas in the awake macaque monkey while the animal performed a binocular rivalry task. In primary visual cortex (V1), only a small fraction of cells weakly modulate their response as a function of the percept of the monkey. The majority of cells responded to one or the other retinal stimulus with little regard to what the animal perceived at the time. In contrast, in a high-level cortical area such as the inferior temporal (IT) cortex along the ventral pathway, almost all neurons responded only to the perceptual dominant stimulus, that is, to the stimulus that was being reported. For example, when a face and a more abstract design were presented, one of these to each eye, a "face" cell fired only when the animal indicated by its performance that it saw the face and not the pattern presented to the other eye (Fig. 51.4). This result implies that the NCC involves activity in neurons in inferior temporal cortex.

In a related perceptual phenomena, *flash suppression*, the percept associated with an image projected into one eye is suppressed by flashing another image into the other

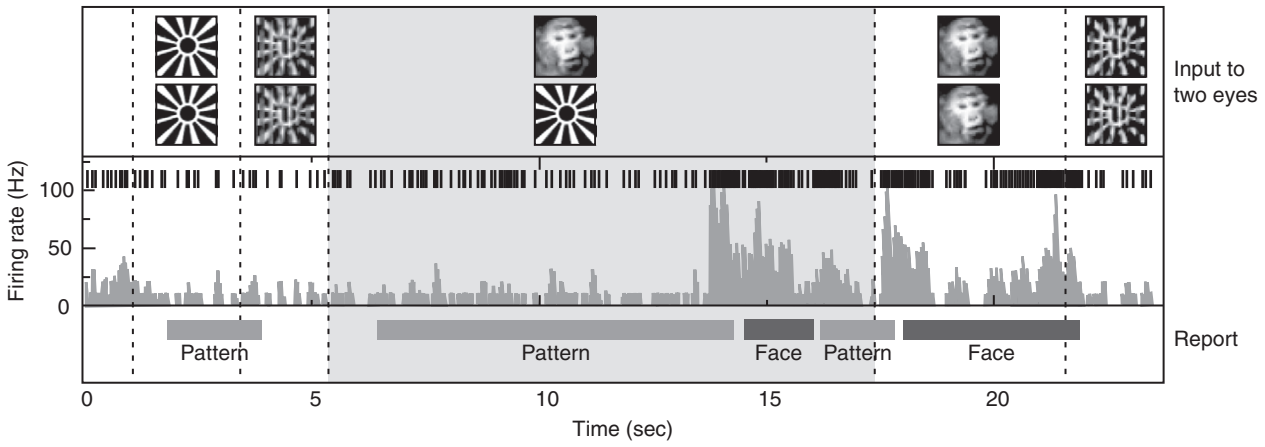


FIGURE 51.4 A fraction of a minute in the life of a typical IT cell while a monkey experiences binocular rivalry. The upper row indicates the visual input to the two eyes, with dotted vertical boundaries marking stimulus transitions. The second row shows the individual spikes, the third the smoothed firing rate, and the bottom row the monkey's behavior. The animal was taught to press a lever when it saw either one or the other image, but not both. The cell responded only weakly to either the sunburst design or to its optical superposition with the image of a monkey's face. During binocular rivalry (gray zone), the monkey's perception vacillated back and forth between seeing the face ("Face") and seeing the bursting sun ("Pattern"). Perception of the face was consistently accompanied (and preceded) by a strong increase in firing rate. *From N. Logothetis (private communication) as modified by Koch (2004).*

eye (while the original image remains). Its methodological advantage over binocular rivalry is that the timing of the perceptual transition is determined by an external trigger rather than by an internal event. The majority of responsive cells in inferior temporal cortex and in the superior temporal sulcus follow the monkey's behavior—and therefore its percept. That is, when the animal perceives a cell's preferred stimulus, the neuron fires; when the stimulus is present on the retina but is perceptually suppressed, the cell falls silent, even though legions of V1 neurons fire vigorously to the same stimulus (Sheinberg & Logothetis, 1997). Single neuron recordings in the medial temporal lobe of epileptic patients during flash suppression likewise demonstrate abolition of their responses when their preferred stimulus is present on the retina but not consciously seen by the patient (Kreiman, Fried, & Koch, 2002).

A large number of fMRI experiments have exploited binocular rivalry and related illusions to identify the hemodynamic activity underlying visual consciousness in humans. They demonstrate quite conclusively that BOLD activity in the upper stages of the ventral pathway (e.g., the fusiform face area and the parahippocampal place area) follow the percept and not the retinal stimulus (Rees & Frith, 2007).

There is a lively debate about the extent to which neurons in primary visual cortex simply encode the visual stimulus or are directly responsible for expressing the subject's conscious percept. That is, is V1 part of the NCC (Crick & Koch, 1995)? It is clear that retinal neurons are not part of the NCC for visual experiences. While retinal neurons often correlate with visual experience, the spiking activity of retinal ganglion cells does not accord with visual experience (for example, there are no photoreceptors at the blind spot; yet no hole in the field of view is apparent; in dreams vivid imagery occurs despite closed eyes and so on). A number of compelling observations link perception with fMRI BOLD activity in human V1 and even LGN (Lee, Blake, & Heeger, 2005; Tong, Nakayama, Vaughan, & Kanwisher, 1998). These data are at odds with single neuron recordings from the monkey. A plausible explanation is that V1 neurons receive feedback signals during conscious perception that are not, however, sufficient to drive the cells to spike (Maier et al., 2008).

Haynes and Rees (2005) exploited multivariate decoding techniques to read out perceptually suppressed information (the orientation of a masked stimulus) from V1 BOLD activity, even though the stimulus orientation was so efficiently masked that subjects performed at chance levels when trying to guess the orientation. That is, although subjects did not give any behavioral indication that they saw the orientation of the stimulus, its slant could be predicted on a single trial basis with better than chance odds from the V1 (but not from V2 or V3) BOLD signal. This finding supports the hypothesis that information present in V1 is not accessible to visual consciousness.

What all of these experiments demonstrate is that the neuronal basis of consciously experienced perceptions can be identified and studied using standard neuroscience tools.

PERCEPTUAL PUZZLES OF CONTEMPORARY INTEREST

The attributes of even simple percepts vary along a continuum. For instance, a patch of color has a brightness and a hue that are variable, just as a simple tone has an associated loudness and pitch. However, is it possible that each particular, consciously experienced, percept is all-or-none? Might a pure tone of a particular pitch and loudness be experienced as an atom of perception, either heard or not, rather than gradually emerging from the noisy background? The perception of the world around us would then be a superposition of many elementary, binary percepts (Sergent & Dehaene, 2004).

Is perception continuous, like a river, or does it consist of a series of discontinuous batches, rather like the discrete frames in a movie (Purves, Paydarfar, & Andrews, 1996; VanRullen & Koch, 2003a). In *cinematographic vision* (Sacks, 2004), a rare form of visual migraine, the subject sees the movement of objects as fractured in time, as a succession of different configurations and positions, without any movement in between. The hypothesis that visual perception is quantized in discrete batches of variable duration, most often related to EEG rhythms in various frequency ranges (from theta to beta), is an old one. This idea is being revisited in light of the discrepancies of timing of perceptual events within and across different sensory modalities. For instance, even though a change in the color of an object occurs simultaneous with a change in its direction of motion, it may not be perceived that way (Bartels & Zeki, 2006; Stetson, Cui, Montague, & Eagleman, 2006),

What is the relationship between endogenous, top-down attention and consciousness? Although these are frequently coextensive—subjects are usually conscious of what they attend to—there is considerable evidence that argues that these are distinct neurobiological processes (Koch & Tsuchiya, 2007). This question is receiving renewed "attention" due to the development of a powerful visual masking technique that can suppress stimuli for minutes at end (*continuous flash suppression*; Tsuchiya & Koch, 2005). Indeed, it has been shown that attention can be allocated to a perceptually invisible stimulus (Naccache, Blandin, & Dehaene, 2002; for a review of many of these experiments see van Boxtel, Tsuchiya, & Koch, 2010a).

That is, subjects can pay attention to something without seeing what they are paying attention to! In other experiments, visual attention has an opposing effect on afterimage duration than visual consciousness (van Boxtel, Tsuchiya, & Koch, 2010b; Watanabe et al., 2011), dissociate selective, visual attention and consciousness—as assayed via stimulus visibility—in human primary visual cortex. Whether or not subjects see what they are looking at makes very little difference to the hemodynamic signal in V1, while attention strongly modulates the BOLD response. When exploring the neural basis of these processes, it is therefore critical to not confound consciousness with attention. Whether the converse also occurs—being conscious of something without attending to it—remains an open question.

FORWARD VERSUS FEEDBACK PROJECTIONS

Many actions in response to sensory inputs are fast, transient, with a limited behavioral repertoire, and unconscious (Milner & Goodale, 1995). Part of the habit system that controls automatic behavior (Shiffrin & Schneider, 1977), these behaviors can be thought of as cortical reflexes. They are characterized by rapid and stereotyped sensory-motor actions called *zombie behaviors* (Koch & Crick, 2001). Examples include eye and hand movements; limb adjustments; keyboard typing; driving a bicycle; executing well-rehearsed sequences in tennis, soccer, or any other athletic activity; and so on. Such automatic behaviors are complemented by a nonhabit, slower, all-purpose conscious mode that deals with more complex aspects of the sensory input (or a reflection of these, as in imagery) and takes time to decide on appropriate responses. A consciousness mode is needed because otherwise a vast number of different zombie modes would be required to react to unusual events. The conscious system may interfere somewhat with the concurrent zombie systems (Logan & Crump, 2009): focusing consciousness onto the smooth execution of a complex, rapid, and highly trained sensory-motor task—dribbling a soccer ball or typing on a keyboard—can interfere with its smooth execution, something well known to athletes and their trainers. Having both a zombie mode that responds in a well-rehearsed and stereotyped manner as well as a slower system that allows time for thinking and planning more complex behavior is a great evolutionary discovery. This latter aspect, planning, may be one of the principal functions of consciousness.

It seems possible that visual zombie modes in the cortex mainly use the dorsal stream in the parietal region (Milner & Goodale, 1995). However, parietal activity can affect consciousness by producing attentional effects on the ventral stream, at least under some circumstances. The basis of this inference is clinical case studies and fMRI experiments in normal subjects (Corbetta & Shulman, 2002). The conscious mode for vision depends largely on the early visual areas (beyond V1) and especially on the ventral "what" stream.

Seemingly complex visual processing (such as detecting animals in natural, cluttered images) can be accomplished by cortex within 130–150 msec (Thorpe, Fize, & Marlot, 1996; VanRullen & Koch, 2003b), far too slow for conscious perception to be involved. It is plausible that such behaviors are mediated by a purely feed-forward moving wave of spiking activity that passes from the retina through V1, into V4, IT and prefrontal cortex, until it affects motorneurons in the spinal cord that control the finger press (as in a typical laboratory experiment). The hypothesis that the basic processing of information is feedforward is supported most directly by the short times required for a selective response to appear in IT cells (Perrett et al., 1992). Indeed, Hung, Kreiman, Poggio, and DiCarlo (2005) were able to decode from the spiking activity of a couple of hundred neurons in monkey IT (over intervals as short as 12.5 msec and about 100 msec after image onset) the category, identity, and even position of a single image flashed onto the retina of the fixating animal. Coupled with a suitable motor output, such a feed-forward network implements a zombie behavior in the Koch and Crick (2001) sense—rapidly and efficiently subserving one task, here distinguishing animal from nonanimal pictures, in the absence of any conscious experience.

Conscious perception requires more sustained, reverberatory neural activity, most likely via global feedback from frontal regions of neocortex back to sensory cortical areas in the back (Crick & Koch, 1995; Lamme & Roelfsema, 2000; Fahrenfort, Scholte, & Lamme, 2008). Such feedback loops would explain why in backward masking a second stimulus, flashed 80–100 msec after onset of a first image, can still interfere with (mask) the percept of the first image. The reverberatory activity builds up over time until it exceeds a critical threshold. At this point, the sustained neural activity rapidly propagates to parietal, prefrontal, and anterior cingulate cortical regions; thalamus, claustrum, and related structures that support short-term memory; and multimodality integration, planning, speech, and other processes intimately related to consciousness. Dehaene et al. (2001) contrasted brain activity during presentation of flashed, masked words with the activity evoked when the words were not masked (Fig. 51.5). BOLD activity in the fusiform (visual) cortex and in a spatially extended network of midline prefrontal and inferior parietal cortices tracked conscious perception.

This is the hypothesis at the heart of the *global workspace* model of consciousness (Baars, 1989; Dehaene & Changeux, 2011). Sensory information is processed in

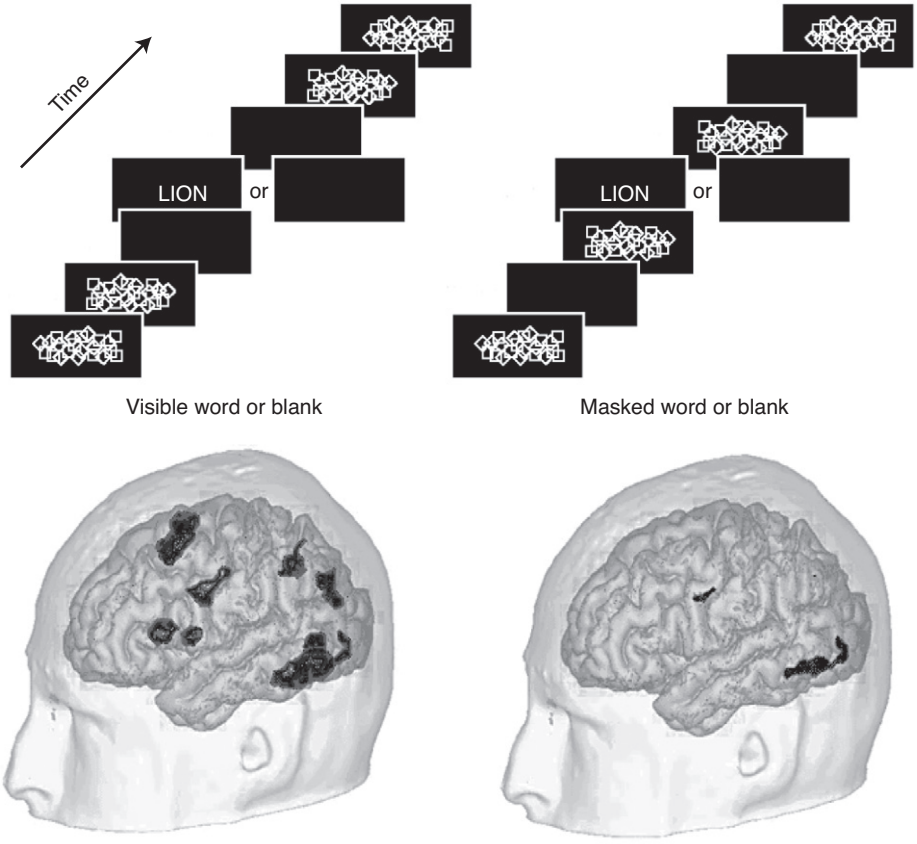


FIGURE 51.5 The Effect of Visual Masking. The brain's response to seen and unseen words. Volunteers looked at a stream of briefly flashed images. No words were seen in the right stream, since each word was preceded and followed by a slide covered by random symbols, perceptually obscuring the letters. This had a dramatic effect on the cortical fMRI response (the activity following the image sequences with words was compared to the sequences with blanks). Both seen and masked words activated regions in the left ventral pathway, but of much different amplitude. Conscious perception triggered additional widespread activation in left parietal and prefrontal cortices. *Modified from Dehaene et al. (2001).*

an array of feed-forward structures, such as primary sensory cortices. If it is powerful enough, it enters into a *global neuronal workspace* located in prefrontal cortical structures. From there it is widely broadcast by long-range cortico-cortical projections and placed into context—for instance, by accessing memory and planning systems. This interpretation feeds back to the sensory representation in visual cortex. Competition inside the global neuronal workspace prevents more than one or a very small number of percepts to be simultaneously and actively represented.

In summary, while rapid but transient neural activity in the thalamo-cortical system can mediate complex behavior without conscious sensation, it is surmised that consciousness requires sustained but well-organized neural activity dependent on integration across cortex.

One hitherto neglected structure in this regard is the claustrum. It is a thin (*ca.* 1 mm) sheet of neurons underneath the insular cortex, extending all the way into temporal, parietal, and frontal lobes, and above the putamen, buried in white matter structures. It projects to, and receives input from, perhaps all cortical regions. Crick and Koch (2005) surmised that the claustrum is critical to the integration of information underlying consciousness, akin to a conductor of the cortical orchestra.

AN INFORMATION-THEORETICAL THEORY OF CONSCIOUSNESS

At present, it is not known to what extent animals whose nervous systems have an architecture considerably different from the mammalian neocortex are conscious. Furthermore, whether artificial systems, such as computers, robots, or the World Wide Web as a whole, which behave with considerable intelligence, are or can become conscious (as widely assumed in science fiction; e.g., the paranoid computer *HAL* in the film *2001*), remains speculative (Koch and Tononi, 2011). What is needed is a theory of consciousness that explains in quantitative terms what type of systems, with what architecture, can possess conscious states.

Progress in the study of the NCC on the one hand, and of the neural correlates of nonconscious habit or zombie behaviors on the other, will hopefully lead to a better understanding of what distinguishes neural structures or processes that are associated with consciousness from those that are not. Yet such an opportunistic, data-driven approach will not lead to an understanding of why certain structures and processes have a privileged relationship with subjective experience. For example, why is it that neurons in corticothalamic circuits are essential for

conscious experience, whereas cerebellar neurons, despite their huge numbers, are not? And what is wrong with cortical zombie systems that make them unsuitable to yield subjective experience? Or why is it that consciousness wanes during slow-wave sleep early in the night, despite levels of neural firing in the thalamocortical system that are comparable to the levels of firing in wakefulness?

Information theory may be such a tool that establishes at the fundamental level what consciousness is, how it can be measured, and what requisites a physical system must satisfy in order to generate it (Chalmers, 1996; Tononi & Edelman, 1998).

The only promising candidate for such a theoretical framework is the *information integration theory of consciousness* (Tononi, 2008). It posits that consciousness is extraordinarily *informative*. Any one particular conscious state rules out a huge number of alternative experiences. Classically, the reduction of uncertainty among alternatives constitutes information. For example, when a subject consciously experiences reading this particular phrase, a huge number of other possible experiences are ruled out (consider all possible written phrases that could have been written in this space, in all possible fonts, ink colors, and sizes; think of the same phrases spoken aloud, or read and spoken, and so on). Thus, every experience represents one particular conscious state out of a huge repertoire of possible conscious states.

Furthermore, information associated with the occurrence of a conscious state is also *integrated*. Any one experience is an indivisible whole. It cannot be subdivided into components that are experienced independently. Again, the conscious experience of this particular phrase cannot be experienced as subdivided into, say, the conscious experience of nouns independently of the verbs. Similarly, visual shapes cannot be experienced independently of their color, nor can the left half of the visual field of view be experienced independently of the right half.

Based on these two considerations, the theory claims that a physical system can generate consciousness to the extent that it integrates information. This requires that the system has a large repertoire of available states (information) that cannot be decomposed into a collection of causally independent subsystems (integration).

Importantly, the theory introduces a measure of a system's capacity to integrate information. This measure, called Φ, is obtained by determining the minimum repertoire of different states that can be produced in one part of the system by perturbations of its other parts (Tononi, 2008). Φ can loosely be thought of as the representational capacity of the system (as in Fig. 51.1). Although Φ is not easy to calculate exactly for realistic systems (Balduzzi & Tononi, 2008), it can be estimated. Thus, by using simple computer simulations, it is possible to show that Φ is high for neural architectures that conjoin functional specialization with functional integration, like the mammalian thalamocortical system. Conversely, Φ is low for systems that are made up of small, quasi-independent modules, like the cerebellum, or for networks of randomly or uniformly connected units.

The notion that consciousness has to do with the brain's ability to integrate information has been tested directly by transcranial magnetic stimulation (TMS). A coil is placed above the skull, and a brief and intense magnetic field generates a weak electrical current in the gray matter underlying the skull. Massimini et al. (2005) compared multichannel EEG of awake and conscious subjects in response to TMS pulses to the EEG when the same subjects were deeply asleep—a time during which consciousness is much reduced. During quiet wakefulness, an initial response at the stimulation site was followed by a sequence of waves that moved to connected cortical areas several centimeters away. During slow-wave sleep, by contrast, the initial response was stronger but was rapidly extinguished and did not propagate beyond the stimulation site. Thus, the fading of consciousness during certain stages of sleep, as well as during anesthesia (Alkire, Hudetz, & Tononi, 2008) and during loss of consciousness in patients (Rosanova et al., 2012), may be related, as predicted by the theory, to the breakdown of information integration among specialized thalamocortical modules.

CONCLUSION

Ever since the Greeks first considered the mind-body problem more than two millennia ago, it has been the domain of armchair speculations and esoteric debates with no apparent resolution. Yet many aspects of this ancient set of questions now fall squarely within the domain of science.

In order to advance the resolution of these and similar questions, it will be imperative to record from a large number of neurons simultaneously at many locations throughout the cortico-thalamo system and related satellites in people or behaving animals, such as monkeys or mice. Progress in understanding the circuitry of consciousness therefore demands a battery of behaviors (akin to but different from the well-known Turing test for intelligence) that the subject—a newborn infant, immobilized patient, monkey, or mouse—has to pass before considering him, her, or it to possess some measure of conscious perception. This is not an insurmountable step for mammals such as the monkey or the mouse that share many behaviors and brain structures with humans. For example, one particular mouse model of contingency awareness (Han et al., 2003) is based on the differential requirement for awareness of trace versus delay associative eyeblink conditioning in humans (Clark & Squire, 1998).

The growing ability of neuroscientists to manipulate in a reversible, transient, deliberate, and delicate manner

genetically identified populations of neurons using methods from molecular biology (Gradinaru, Mogri, Thompson, Henderson, & Deisseroth, 2009; Lin et al., 2011) opens the possibility of moving from correlation—observing that a particular conscious state is associated with some neural or hemodynamic activity—to causation. Exploiting these increasingly powerful tools depends on the simultaneous development of appropriate behavioral assays and model organisms amenable to large-scale genomic analysis and manipulation, particularly in mice (Lein et al., 2007). For example, if the masking experiment of Dehaene and colleagues (Fig. 51.5) is repeated using appropriately adjusted parameters in the mouse's visual system, all cortico-cortical feedback can be transiently inactivated to test to what extent this massive and sustained activity that in humans is associated with conscious perception depends on top-down connections.

It is the combination of such fine-grained neuronal analysis in mice and monkeys, ever more sensitive psychophysical and brain imaging techniques in patients and healthy individuals, and the development of a robust theoretical framework that lend hope to the belief that human ingenuity can, ultimately, understand in a rational manner one of the central mysteries of life.

References

Allman, J. M. (1999). *Evolving brains*. New York: Scientific American Library.

Alkire, A. M., Hudetz, A. C., & Tononi, G. (2008). Consciousness and anesthesia. *Science, 322,* 876–880.

Baars, B. J. (1989). *A cognitive theory of consciousness*. Cambridge, UK: Cambridge University Press.

Balduzzi, D., & Tononi, G. (2008). Integrated information in discrete dynamical systems: motivation and theoretical framework. *PLoS Computational Biology, 4,* e1000091.

Bartels, A., & Zeki, S. (2006). The temporal order of binding visual attributes. *Vision Research, 46,* 2280–2286.

Blake, R., & Logothetis, N. K. (2002). Visual competition. *Nature Reviews Neuroscience, 3,* 13–21.

Bogen, J. E. (1995). On the neurophysiology of consciousness: I. An Overview. *Consciousness & Cognition, 4,* 52–62.

Boly, M., Garrido, M. I., Gosseries, O., Bruno, M. A., Boveroux, P., Schnakers, C., … Friston, K. (2011). Preserved feedforward but impaired top-down processes in the vegetative state. *Science, 332,* 858–862.

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. New York: Oxford University Press.

Chalmers, D. J. (2000). What is a neural correlate of consciousness?. In T. Metzinger (Ed.), *Neural correlates of consciousness: Empirical and conceptual questions* (pp. 17–40). Cambridge: MIT Press.

Clark, R. E., & Squire, L. R. (1998). Classical conditioning and brain systems: The role of awareness. *Science, 280,* 77–81.

Collini, E., Wong, C. Y., Wilk, K. E., Curmi, P. M. G., Brumer, P., & Schoes, G. D. (2010). Coherently wired light-harvesting in photosynthetic marine algae at ambient temperature. *Nature, 463,* 644–648.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience, 3,* 201–215.

Crick, F. C., & Koch, C. (1995). Are we aware of neural activity in primary visual cortex? *Nature, 375,* 121–123.

Crick, F. C., & Koch, C. (2003). A framework for consciousness. *Nature Neuroscience, 6,* 119–127.

Crick, F. C., & Koch, C. (2005). What is the function of the claustrum. *Philosophical Transactions of the Royal Society of London Series B, 360,* 1271–1279.

Dehaene, S., & Changeux, J-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron, 70,* 200–227.

Dehaene, S., Naccache, L., Cohen, L., Bihan, D. L., Mangin, J. F., Poline, J. B., & Riviere, D. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience, 4,* 752–758.

Desmurget, M., Reilly, K. T., Richard, N., Szathmari, A., Mottolese, C., & Sirigu, A. (2009). Movement intention after parietal cortex stimulations in humans. *Science, 324,* 811–813.

Edelman, D. B., Baars, J. B., & Seth, A. K. (2005). Identifying hallmarks of consciousness in non-mammalian species. *Consciousness & Cognition, 14,* 169–187.

Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. F. (2008). The spatiotemporal profile of cortical processing leading up to visual perception. *Journal of Vision, 8*(1), 12.1–12.

Giurfa, M., Zhang, S., Jenett, A., Menzel, R., & Srinivasan, M. V. (2001). The concepts of "sameness" and "difference" in an insect. *Nature, 410,* 930–933.

Gradinaru, V., Mogri, M., Thompson, K. R., Henderson, J. M., & Deisseroth, K. (2009). Optical deconstruction of Parkinsonian neural circuitry. *Science, 324,* 354–359.

Haggard, P., & Eimer, M. (1999). On the relation between brain potentials and conscious awareness. *Experimental Brain Research, 126,* 128–133.

Haggard, P. (2008). Human volition: Towards a neuroscience of will. *Nature Reviews Neuroscience, 9,* 934–946.

Han, C. J., O'Tuathaigh, C. M., van Trigt, L., Quinn, J. J., Fanselow, M. S., Mongeau, R, … Anderson, D. J. (2003). Trace but not delay fear conditioning requires attention and the anterior cingulate cortex. *Proceeding of the National Academy of Sciences of the United States of America, 100,* 13087–13092.

Haynes, J-D., & Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neuroscience, 8,* 686–691.

Herculano-Houzel, S. (2009). The human brain in numbers: A linearly scaled-up primate brain. *Frontiers in Human Neuroscience, 3*(31), 1.

Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science, 310,* 863–866.

Jeon, D., Kim, S., Chetana, M., Jo, D., Ruley, H. E., Lin, S-Y., … Shin, H-S. (2010). Observational fear learning involves affective pain system and Ca_V 1.2 Ca^{2+} channels in ACC. *Nature Neuroscience, 13,* 482–488.

Kim, C-Y., & Blake, R. (2005). Psychophysical magic: Rendering the visible "invisible." *Trends in Cognitive Sciences, 9,* 381–388.

Koch, C. (2004). *The quest for consciousness: a neurobiological approach*. Denver, CO: Roberts.

Koch, C. (2012). *Consciousness: Confessions of a romantic reductionist*. Cambridge, MA: MIT Press.

Koch, C., & Crick, F. C. (2001). On the zombie within. *Nature, 411,* 893.

Koch, C., & Hepp, K. (2010). The relation between quantum mechanics and higher brain functions: Lessons from quantum computation and neurobiology. In R. Y. Chiao, et al. (Eds.), *Amazing light: New light on physics, cosmology and consciousness* (pp. 584–600). Cambridge, UK: Cambridge University Press.

Koch, C., & Reid, C. R. (2012). Observatories of the mind. *Nature, 483,* 397–398.

Koch, C., & Tononi, G. (2011 June). A turing test for consciousness. *Scientific American, 304,* 44–47.

Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: Two distinct brain processes. *Trends in Cognitive Sciences, 11,* 16–22.

Kreiman, G., Fried, I., & Koch, C. (2002). Single-neuron correlates of subjective vision in the human medial temporal lobe. *Proceeding of the National Academy of Sciences of the United States of America, 99,* 8378–8383.

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences, 23,* 571–579.

Laureys, S. (2005). The neural correlate of (un)awareness: Lessons from the vegetative state. *Trends in Cognitive Sciences, 9,* 556–559.

Lee, S. H., Blake, R., & Heeger, D. J. (2005). Traveling waves of activity in primary visual cortex. *Nature Neuroscience, 8,* 22–23.

Lein, E. S., Hawrylycz, M. J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., … Jones, A. R. (2007). Genome-wide atlas of gene expression in the adult mouse brain. *Nature, 445,* 168–176.

Lemon, R. N., & Edgley, S. A. (2010). Life without a cerebellum. *Brain, 133,* 649–654.

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): The unconscious initiation of a freely voluntary act. *Brain, 106,* 623–642.

Lin, D., Boyle, M. P., Dollar, P., Lee, H., Lein, E. S., Perona, P., & Anderson, D. J. (2011). Functional identification of an aggression locus in the mouse hypothalamus. *Nature, 470,* 221–226.

Logan, G. D., & Crump, M. J. C. (2009). The left hand doesn't know what the right hand is doing: The disruptive effects of attention to the hands in skilled typewriting. *Psychological Science, 20,* 1296–1300.

Logothetis, N. (1998). Single units and conscious vision. *Philosophical Transactions Royal Society of London B, 353,* 1801–1818.

Maier, A., Wilke, M., Aura, C., Zhu, C., Ye, F. Q., & Leopold, D. A. (2008). Divergence of fMRI and neural signals in V1 during perceptual suppression in the awake monkey. *Nature Neuroscience, 11,* 1193–1200.

Massimini, M., Ferrarelli, F., Huber, R., Esser, S. K., Singh, H., & Tononi, G. (2005). Breakdown of cortical effective connectivity during sleep. *Science, 309,* 2228–2232.

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*. Oxford: Oxford University Press.

Monti, M. M., Vanhaudenhuyse, A, Coleman, M. R., Boly, M., Pickard, J. D., Tshibanda, L., … Laureys, S. (2010). Willful modulation of brain activity in disorders of consciousness. *New England Journal of Medicine, 362,* 579–589.

Naccache, L., Blandin, E., & Dehaene, S. (2002). Unconscious masked priming depends on temporal attention. *Psychological Sciences, 13,* 416–424.

Parvizi, J., & Damasio, A. R. (2001). Consciousness and the brainstem. *Cognition, 79,* 135–160.

Purves, D., Paydarfar, J. A., and Andrews, T. J. (1996). The wagon wheel illusion in movies and reality. *Proceeding of the National Academy of Sciences of the United States of America, 93,* 3693–3697.

Rees, G., & Frith, C. (2007). Methodologies for identifying the neural correlates of consciousness. In M. Velmans & S. Schneider (Eds.), *The blackwell companion to consciousness* (pp. 553–566). Oxford: Blackwell.

Rosanova, M., Gosseries, O., Casarotto, S., Boly, M., Casali, A. G., Bruno, M-A., Mariott, M., Boveroux, P., Tononi, G., Laureys, S., & Massimini, M. (2012). Recovery of cortical effective connectivity and recovery of consciousness in vegetative patients. *Brain, 6,* 1–13.

Sacks, O. (2004). In the river of consciousness. *New York Review Books, 51,* 41–44.

Schiff, N. D. (2010). Recovery of consciousness after brain injury. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (4th ed., pp. 1123–1136). Cambridge, MA: MIT Press.

Searle, J. (2004). *Mind: A brief introduction*. New York, NY: Oxford University Press.

Sergent, C., & Dehaene, S. (2004). Is consciousness a gradual phenomenon? Evidence for an all-or-none bifurcation during the attentional blink. *Psychological Science, 15,* 720–728.

Sheinberg, D. L., & Logothetis, N. K. (1997). The role of temporal cortical areas in perceptual organization. *Proceeding of the National Academy of Sciences of the United States of America, 94,* 3408–3413.

Shiffrin, R., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84*(2), 127–190.

Slater, R., Cantarella, A., Gallella, S., Worley, A., Boyd, S., Meek, J., & Fitzgerald, M. (2006). Cortical pain responses in human infants. *Journal of Neuroscience, 26,* 3662–3666.

Stetson, C., Cui, X., Montague, P. R., & Eagleman, D. M. (2006). Motor-sensory recalibration leads to reversal of action and sensation. *Neuron, 51,* 651–659.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381,* 520–522.

Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron, 21,* 753–759.

Tononi, G. (2008). Consciousness as integrated information: A provisional manifesto. *Biological Bulletin, 215,* 216–242.

Tononi, G., & Edelman, G. M. (1998). Consciousness and complexity. *Science, 282,* 1846–1851.

Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nature Neuroscience, 8,* 1096–1101.

VanRullen, R., & Koch, C. (2003a). Is perception discrete or continuous? *Trends in Cognitive Sciences, 7,* 207–213.

VanRullen, R., & Koch, C. (2003b). Visual selective behavior can be triggered by a feed-forward process. *Cognitive Neuroscience, 15,* 209–217.

van Boxtel, J. A., Tsuchiya, N, & Koch, C. (2010a). Consciousness and attention: On sufficiency and necessity. *Frontiers in Consciousness Research, 1,* 1–13.

van Boxtel, J. A., Tsuchiya, N., & Koch, C. (2010b). Opposing effects of attention and consciousness on afterimages. *Proceeding of the National Academy of Sciences of the United States of America, 107,* 8883–8868.

Villablanca, J. R. (2004). Counterpointing the functional role of the forebrain and of the brainstem in the control of the sleep-waking system. *Journal of Sleep Research, 13,* 179–208.

Watanabe, M., Cheng, K., Murayama, Y., Ueno, K., Asamizuyu, T., Tanaka, K., & Logothetis, N. (2011). Attention but not awareness modulates the BOLD signal in the human V1 during binocular suppression. *Science, 334,* 829–831.

Wegner, D. M. (2003). *The illusion of conscious will*. Cambridge, MA: MIT Press.

Christof Koch

This page intentionally left blank

Index

A

AADC, *see* L-Aromatic amino acid decarboxylase
AAT, *see* Aspartate amino transferase
Abnormal breathing patterns, CNS lesions produce, *753–754*, *763*
AC, *see* Adenylate cyclase
Accessory olfactory bulb (AOB), pheromone detection, 527
Accommodation, vergence eye movements and, 699
Accumbens (nucleus accumbens), 873, 877, 879, 883, *884*, 885–886, 888, 891–892, *893*, *see also* Basal ganglia
Acetylcholine (ACh), 117, 655
  biosynthesis, 126–127
  consciousness role, 1095
Acetylcholine receptor (AChR), 385, *see also* Muscarinic acetylcholine receptor; Nicotinic acetylcholine receptor
  α-bungarotoxin inhibition, *448*
  clustering, *388*, 389–390, 392
  expression, *388*, 391–392, 394
    muscle-autonomous
      and neural-dependent patterning, 391–392
    patterning of, 386–387
Acetylcholinesterase (AChE), 127, 658
ACh, *see* Acetylcholine
AChE, *see* Acetylcholinesterase
AChR, *see* Acetylcholine receptor
ACTH, *see* Adrenocorticotropin
Actin
  axonal growth, 374
  cycle, 374
  cytoskeleton role in cell movement and transport, 78–80
  genes, 78
Action potentials
  bursts and calcium low-threshold currents, 114
  calcium-binding proteins and action potential coupling to exocytosis, 153
  generation, 102–105
  initiation, 105
  propagation speed and myelination, 106
  refractory periods, 105–106
Action tremor, *688*
Activator protein 1 (AP-1), transcription, 206–207
Active zones, 386–387, *387*, *389*
Activity-dependent neuromodulation, 1014
Addiction
  brain imaging studies, 891–896
  chronic drug administration compromise of reward systems, 887
  cycle, 881–883
  diagnostic criteria, *881*
  drug reward neurobiological substrates
    alcohol, 886
    basal forebrain, 883
    mesolimbic dopamine system
      cocaine, 883–884
      neural substrates of sensitization, 884–885
      nicotine, 886
    opiates, 885–886
    Δ^9-tetrahydrocannabinol, 886
  economic impact, 879
  epidemiology of drug abuse, 879–881
  negative affect state, 886–888
  negative reinforcement, 887
  neuroadaptation, abstinence and relapse, 889–891
  neurochemical adaptation
    corticotropin-releasing factor, 888–889
    dopamine, 888, 896
    norepinephrine, 889
    serotonin, 888
  signal transduction mechanisms, 892–896
  withdrawal from chronic administration of drugs, 887
  withdrawal neurobiological substrates aversion, 886–888
Adenylate cyclase (AC)
  addiction role, 892
  coincidence detection, 193
  fine-tuning of cAMP, 192
  inhibition, 192
  isoforms of, 192
  olfactory signaling, 523
  receptor coupling, 192–193
  regulation of, 192–193
ADHD, *see* Attention deficit hyperactivity disorder
ADP-ribosylation factors (ARFs), 68
Adrenal medulla, 733
Adrenergic receptors, 183–184
Adrenocortical dysfunction, *794*
Adrenocorticotropin (ACTH), 803
Adult neurogenesis, *352*
Advanced sleep phase syndrome (ASPS), *829*
AE, *see* Anterior endoderm
Aging
  life expectancy trends, 927
  memory effects
    animal models, 928–929
    neurobiology of decline
      hippocampus, 929
      neuron loss, 929–930
      prefrontal cortex, 931–933
      subcortical systems, 933
    pathology, 940–943
    variability, 927–929
Agnosia
  associative agnosia, 948–950, *948–950*
  case study, *949*
  inferior temporal cortex, 948, 950
  types, 948–949
Agouti-related protein (AgRP), food intake regulation, 778
Agrin, 385
  signal for postsynaptic differentiation, 389–390
Agrin-mediated signaling, and synapse formation
  for Lrp4, 390–391
  for MuSK, 390–391
AgRP, *see* Agouti-related protein
AIP, *see* Anterior intraparietal area
Alanine amino transferase (ALAT), 274
Alcohol, addiction mechanisms, 886
Aldosterone, salt appetite regulation, 793–795
Allostasis, 889, *891*
Alpha cells, 584
ALS, *see* Amyotrophic lateral sclerosis

Note: Page numbers in *italics* indicate figures, tables and text boxes.

Alzheimer's disease (AD), *1032–1033*
clinical features, 941
epidemiology, 941
genetics, 942–943
neuropathology, 941–942
treatment prospects, 943
Amacrine cell, *232*, 582–583
AME, *see* Axial mesendoderm
γ-Aminobutyric acid (GABA), 655–656
biosynthesis, 128
inactivation, 128
receptors, *see* $GABA_A$ receptor; $GABA_B$ receptor
sleep control, 855–856, 858
storage and release, 128
transmitter *versus* metabolic pools, 128
Amino-3-hydroxy-5-methylisoxazolproprionic acid (AMPA) receptor, 171, 239, *252*
dendrite development role, *470*, 470–472
Amino-terminal domain (ATD), 173, *397*, 399
Amnesia
anterograde amnesia animal models, 1033–1035, 1039
patient H.M. and, 1031–1033
retrograde amnesia and consolidation, 1039–1040
AMPA receptor, *see* Amino-3-hydroxy-5-methylisoxazolproprionic acid receptor
Amygdala memory system
central nucleus, 1044
emotional disposition acquisition, 1046
fear conditioning, 1044, *1045*
modulation, 1046–1047
motivation role, 879
Amyloid β-protein (Aβ), 941
Amyloid precursor protein (APP), *291*, 432, 942, *1032*
Amyotrophic lateral sclerosis (ALS), 385, *622*
Androgen-response elements (AREs), 811
Animal communication, 1053–1057, *see also* Communication
ANP, *see* Atrial natriuretic peptide
ANS, *see* Autonomic nervous system
Anterior endoderm (AE), 293
Anterior intraparietal area (AIP), 642
Anterior neural plate (ANB), 294
Anterior neural ridge (ANR), 300
Anterior nidopallium (LMAN), birdsong learning neural pathway, 483
Anterior visceral endoderm (AVE), 294
Anterograde amnesia, animal models, 1033–1035, 1039
Anterograde organizers, 396
Anterograde signals, *397*, 399
postsynaptic neurotransmitter receptors, 399
Anteroposterior (AP) axis, 288
homeobox gene expression, 298–299
local organizers of, 300–301
molecular basis, 293–294
spinal nerves, 301–302
Anteroposterior patterning, *see* Neural patterning
Anthropoids, brains of, 913–914
Anticonvulsant medications, 1031
Anti-diuretic hormone (ADH), 813
AOB, *see* Accessory olfactory bulb
AP, *see* Area postrema
AP-1, *see* Activator protein 1
AP axis, *see* Anteroposterior axis
Aplysia ubiquitin hydrolase (ApUch), 1013
Apneustic breathing, *753*
Apolipoprotein E (ApoE), alleles, 942
Apoptosis, 412, *see also* Programmed cell death
mechanisms, 412–416
APP, *see* Amyloid precursor protein
Appetite, *see* Food intake
Aquaporins, 785, *786*
Arachnoid, 35–36
Arcuate nucleus (ARC), 777–779, *778*
Area postrema (AP), 721, *722*, 725–726
ARFs, *see* ADP-ribosylation factors
Arginine vasopressin (AVP), 813–814
circadian control, 831
osmotic homeostasis regulation, 785
secretion signals, 790–792
L-Aromatic amino acid decarboxylase (AADC), catecholamine biosynthesis, 122
Arousal of consciousness, 1093–1095
Arterial chemoreceptor, 761
Aspartate amino transferase (AAT), 274
Aspartate, excitatory neurotransmission, 128–129
ASPS, *see* Advanced sleep phase syndrome
Associative learning, 1009–1010, 1014–1016
Astrocyte, 51–52
astrocyte-neuron metabolic unit, 283–284
glutamate metabolism, 274–276
glycogen metabolism, 282
hepatic encephalopathy role, *275–276*
lactate release, 281–282
positron emission tomography studies of metabolism, 279–281
Asynergia, definition, *688*
Ataxia, definition, *688*
Ataxic breathing, *754*
ATD, *see* Amino-terminal domain
Atrial natriuretic peptide (ANP), volume homeostasis regulation, 789, *793*, 795
Attention, 989–990
competition bias by nonspatial feedback, 1005–1006
filtering of unwanted information, 1004–1005
flexibility, 990
fronto-parietal attention control system, 994–998
lateral geniculate nucleus neural response modulation, 1000–1001
neglect syndrome, 990–992
neuronal receptive fields, 1002–1004
neuronal responses, 998–1000, *999*
search tasks, 1001–1002
stimulus, *1003*
visual cortex neural activation with/without stimulation, *992–993*, 992–994
visual system, processing information about objects, 998–1000
Attention deficit hyperactivity disorder (ADHD), *673*, *1080–1081*
Audition
and auditory cortex, 572
complex signal coding, 573–574
and sound localization, 573
and auditory nerve
descending systems to hair cells and nerve fibers, 564–565
frequency response, 562–563
masking, 564
neuron types, 562
phase locking of responses, 563
sound level and spontaneous activity response, 563
central nervous system
contralateral *versus* ipsilateral ear response, 568–569
interaural level difference, 484
interaural time difference, 484–486
poststimulus time histogram, *565*, 566–567
response map, 567, *569*
tonotopic organization of pathways, 566
cochlea
endolymph, 559
hair cells, 554–555
implants, *561*
outer hair cell mechanical properties and amplification, 557–559
receptor potentials from inner hair cells, 556–557, *558*
sensorineural hearing loss, 559
stereocilia channels and transducer gating, 555–556
structure, 554, *555*
ear, 553
echolocation in bats, *573–574*
endbulbs and calyces, *568*
inferior colliculus, 570–572
interaural level difference, *569*, 569–570
interaural time difference, *569*, 569–570
loss, sensorineural, 559
medial geniculate, 572
otoacoustic emissions, *560*
Auditory cortex, *see* Audition
Auditory nerve fibers
descending systems to hair cells and nerve fibers, 564–565
frequency response, 562–563

masking, 564
neuron types, 562
phase locking of responses, 563
sound level and spontaneous activity response, 563
Augmentation, 158–160, *159*
Autism, *398*, 934
genetics, 936–938
neuropathology, 934–936
Autonomic, 849, *see also* Sleep
Autonomic afferents, 740–743
Autonomic function, circadian regulation, 840–841
Autonomic motor pattern, *730*
Autonomic nervous system (ANS), 750, *see also* Enteric nervous system; Parasympathetic nervous system; Sympathetic nervous system
central integration in hypothalamus, 724
divisions, 33
enteric division of, 736–738
functional overview, 729–730
future of, 746–747
hierarchically organized circuits in CNS, 743
affective responses, *745*
lesion and stimulation studies, 744
mapping, 745–746
homeostasis and visceral afferents, 740–743
pharmacology, 738–740
Autoreceptors, 119, 160
AVE, *see* Anterior visceral endoderm
Aversion, opponent motivational processes, 887–888
Avian embryo, *340*
AVP, *see* Arginine vasopressin
Axial mesendoderm (AME), 293
Axial resistance, 101
Axoaxonic synapses, 4
Axodendritic synapse, 4
Axon, *65*
action potential initiation, 105
Hillock, 42, 62
pathfinding, *see* Growth cone guidance
structure, 62, 64
topographic map, *see* Topographic map
Axon reflexes, 742
Axonal transport, 86–89
neurofilament proteins, 87
regulation, 89–91
Axon-target interaction, initiation of, 395–396
Axosomatic synapse, 4

B

BAC spikes, 247, *250*
simultaneous somatic current injection, *250*
Balance, *see* Posture control
Balint's syndrome, 970
BAR proteins, 154
Bariatric surgery, *775*
Baroreceptor reflex regulation, 761–762
Basal forebrain (BF) nuclei, 851
Basal ganglia
anatomy of, 653–661, *654*
connection, *654*, 655
globus pallidus
damage effects, 665
external segment connectivity, 660
internal segment output, 659–660
movement response and spontaneous activity, 662–663
motor control, 606–607, 668–669
nonmotor functions, 669–675
prefrontal cortex
connectivity, 1070
function comparison, 1087
saccade regulation, 711
signaling in, 661–664
striatum
cortical input, 655
damage effects, 664
neurons, 655
spontaneous activity, 661–662
striatum memory system, *see* Striatum
substantia nigra
damage effects, 665–667
dopaminergic neurons, 48–49
output for eye movements, 659
pars compacta dopamine input to striatum, 660
spontaneous activity, 662–663
subthalamic nucleus
damage effects, 664–665
inputs from frontal lobe, 658
spontaneous activity, 662
ventral system of, *670*
Basal metabolism, hypothalamic control of, 804–806
Basal pontine nuclei, 706
Basal sympathetic tone, 759–760
Basilar artery, 36
Basket cell, *681*, 683–684
Basolateral membrane shape receptor potentials, 556–557
Bat echolocation, *573–574*
BCIs, *see* Brain–computer interfaces
Bcl-2, apoptosis role, 413
BDNF, *see* Brain–derived neurotrophic factor
Behavioral neuroscience, *874*
Beta cells, 584
Bicuculline, $GABA_A$ binding, 170
Bidirectional organizers, 396
Bidirectional synapse organizers, transsynaptic adhesion complexes in, 397–399
Binocular rivalry, 1097
Binocular vision, 487–489
Bipolar cell, 581
Birdsong learning, 481–483
hormonal regulation, 483
neural pathway, 483
sensitive period for memorization, 482
vocal learning, 482–483
Blood pressure, 783–784
Blood volume, 783–784
Blood-brain barrier, function, 7, 54, 56
Blood-oxygen-level-dependent (BOLD), *266*
signal, 859
BMIs, *see* Brain-machine interfaces
BMPs, *see* Bone morphogenetic proteins
Body mass index (BMI), *774*
Bone morphogenetic proteins (BMPs)
growth cone guidance, 372
neural induction, 288–290, *310*, 319
signaling pathway, *290*
Bottom-up mechanisms, context of stimuli, 1004
Brain, *see also specific components*
blood–brain barrier, 7
computer analogy, 15
cranial nerves, 34, *34*
damage response, 7
development
cerebral metabolic rate changes, 922
cognitive development declarative memory, 924–925
face processing, 925–926
language, 926–927
pathology, 934–940
cortical thickening, 919–920
dendritic spine number, 920–921
dorsolateral prefrontal cortex, 922–923
maturation, 919
prefrontal cortex and visual search, 926
synapse abundance, 921
visual system
dorsal visual stream and spatial attention, 924
functional development, 923–924
ventral visual stream and object representation, 924
white matter changes, 922
evolution, mammals, 901–906, *903*
neocortex origins, 901–904, *904*, 906
primates
early anthropoids, 913–914
early primates, *907*, 909–913
evolution, 906–909
hominin brain, 914–916
prosimian, 906, *907*
size of, *910*
tools for study, *902*
glia support of neurons, 7
imaging studies, 891–896
molecules transport in, 56–57
neuronal circuitry, 6–7
principles of organization and function, 6
reinforcement systems, *878*
reward systems, 887
sexual differentiation of, 815
steroid hormone receptors, 815
stimulation reward, *878*
stress systems, 888–889

Brain (*Cont.*)
structures, related to language, 1060–1061
vascular, 36–37
vasculature, 55–57
in disease states, 57
Brain energy metabolism
astrocytes
astrocyte-neuron metabolic unit, 283–284
glutamate metabolism, 274–276
glycogen metabolism, 282
positron emission tomography, 279
coupling
neuronal activity to blood flow, 264–268
positron emission tomography studies, *266*
endothelial cell, 276–277
glia, 276–277
glucose transporters, 277–278
glucose utilization, 261–262
glycolysis, 269
ketone body utilization, 262–263
lactate utilization, 263
magnetic resonance spectroscopy, 263
neuronal energy-consuming processes, 271–273
pentose phosphate pathway, 269–270
pyruvate utilization, 263
reactive oxygen species generation and glutathione protection, 270–271
tricarboxylic acid cycle, 269
Brain stem
arousal systems of, *851*
eye movement regulation, 705–707
food intake regulation, 780–781
medial postural system, 635
sleep regulatory system, 855–858
spinal cord connections in motor system, 606, 628, 639
Brain–computer interfaces (BCIs), *648*
Brain-derived neurotrophic factor (BDNF)
dendrite growth and branching regulation, 461
discovery, 418
Brain-machine interfaces (BMIs), *648*
Breathing, 749–750
CNS lesions producing abnormal patterns, *753–754*, *763*
controlling, 761–762
disorders of, *763–764*
distinct oscillators, *755*
experimental analysis of, *751*
history of study, *750*, 750–751
mechanoreceptors in lung afferents, 756
Breuer-Hering reflex, 756
C fibers, 757
rapidly adapting pulmonary stretch receptors, 757
slowly adapting pulmonary stretch receptors, 756–757
motor output modulation and plasticity, 762–765
pattern-generating neurons, 751–752, *752*
preBötzinger complex for, 750, *753*, 753–754
respiratory neurons discharge pattern, 752
rhythm-generating neurons, 753–756
Breuer–Hering reflex, 756
Broca's area, *1060*
Bronchopulmonary C fibers, 757
Bulbosacral division, 735
α–Bungarotoxin, acetylcholine receptor inhibition, *448*

C

C fibers, 757
CA1 pyramidal neurons, 240, *240*, *252*
inhibitory interneurons, *250*
K_v4 channels, 244
CA3–CA1 synapse, 1016–1018, *1017*
Cadherins, 400
dendrite growth and branching regulation, 464–466
Cajal, Santiago Ramôn y, biography, *20*
Cajal-retzius cells, 356–357
Calcineurin, structure and regulation, 202–203
Calcium
facilitation, augmentation, and potentiation regulation by residual calcium, 160
G protein-coupled receptor signaling, 195
long-term potentiation induction, 1018–1019
neurotrophin receptor signaling, 423
Calcium current
high-threshold current subtypes in neurons, 114
low-threshold currents and action potential bursts, 114
second messenger systems, 113
Calcium microdomain, vesicle release, 143
Calcium-binding proteins, action potential coupling to exocytosis, 153
Calcium/calmodulin protein kinase II (CaMKII)
cognition role, 201–202
dendrite development role, 471–472
short-term sensitization in *Aplysia*, 1011
Calcium-induced calcium release (CICR), 471
Calmodulin, mediation of calcium effects, 195–196
Caloric homeostasis, 767–768
food intake regulation, *see* Food intake
insulin role, 768–769
Calyce, central auditory processing, *568*
cAMP, *see* Cyclic adenosine monophosphate
CAMs, *see* Cell adhesion molecules
Carbon dioxide, chemoreceptor, 761–762
Cardiovascular sympathetic preganglionic neurons, 759
Cardiovascular system, 750
baroreceptor reflex regulation, 761–762
basal neural tone to, 759
disease, *760*
exercise-induced changes, 764–765
organization and functions of innervation, *757*, 757–759
CART, *see* Cocaine-amphetamine-related transcript
Catecholamine, *see also* Dopamine; Epinephrine; Norepinephrine
biosynthesis of, 120–122
Catechol-*O*-methyltransferase (COMT), 940
function, 124
inhibitors, *125*
β–Catenin, dendrite growth and branching regulation, 464–466
Caudal ventral respiratory group (cVRG), 751
Caudal ventrolateral medulla (CVLM), 761
Caudate neurons, 662
Caveolae, 57
CBF, *see* Cerebral blood flow
CBF neurons, *see* Cholinergic basal forebrain neurons
CCK, *see* Cholecystokinin
Cdk5, dendrite orientation role, 461
CE, *see* Convergent-extension
Celiac ganglion, 732
Cell adhesion molecules (CAMs), 340, 373
Cell assembly, 3, 1029, 1038
Cell fate, 289
regulation, 314–316
Cell motility molecular mechanisms
extracellular matrix, 351
GTPases, 351
Rac and Cdc42 signaling, 352
Cell-autonomous circadian oscillator, 821–824, *822*
Cell–cell contact, 395, *396*
Cellular determination
asymmetric cell division and cell fate, 314–316
cerebral cortex cells, 309, *311*, 332–335
motor neurons, 330–332, *331*, *333*
neural stem cells, *318–319*
neurogenesis, 310–312
neurons and glia, 316–319
peripheral nervous system sensory neurons, 320–323
proneural and neurogenic genes, 312–314, *314*
retina, *see* Retina
spinal cord, *see* Spinal cord
transcriptional hierarchies and networks, 314
Cellular organization of sensory cortex, *509*

Central autonomic control systems, 849
Central nervous system (CNS), 97, 287, 437
  audition
    contralateral *versus* ipsilateral ear response, 568–569
    poststimulus time histogram, *565*, 566–567
    response map, 567–568, *569*
    tonotopic organization of pathways, 566
  hierarchically organized ANS circuits in, 743–746
  lesions producing abnormal breathing patterns, *753–754*
  molecular signals for synaptic development, 396–400
  regionalization of, 296–306
  synapse formation in, 394–401
Central pattern generator (CPG), *604*
  locomotion cycle, 614–615
  models for study, 622–624
  nonlinear properties of interneurons, 626
  rhythmic movements, 605
Central sleep apnea, 763, 862
Central synapses, assembly hypothetical model for, *396*
Cephalic phase, 768, *769*
Cerebellar Purkinje
  cell, *252*
  neurons, *240*, 247
Cerebellar system, 158–161
Cerebellum
  anatomy, 677–684
  basic circuit, 679–683
  cognition, emotion and, *690–691*
  complete circuit, 683–684
  computation, 691
  coordination, timing, and learning, 693–694
  development, 684
  eye movement regulation, 705–707
  and eyelid conditioning, 691–693
  function theories, 690–691
  gross anatomy, 678–679
  memory system
    anatomy, 1043
    eyeblink classical conditioning, 1043–1044
  motor control, 609–610
  neural circuits
    dentate, 689–690
    fastigius, 689
    interpositus, 689
    vestibular nucleus, 689
  neurons properties, 684–685
    climbing fiber synapse, 685
    high firing rates, 685
    mossy fiber glomerulus, 685
  nuclei
    damage effects, *688*
    input, 686
    purkinje cell, 685–686
  phylogenetic development, 677–678
  synapse elimination, 452–453
  synaptic and intrinsic forms, 687
  vestibular *versus* nonvestibular, *682*
Cerebral blood flow (CBF), 261
Cerebral cortex, *see also* Sensory cortex
  cellular determination, 309, *311*, 332–335
  development, *348*
  developmental thickening, 919–920
  histogenesis, 348–350
  inhibitory interneurons, 45–47
  lobes, 34
  memory system
    anatomy, 1047
    motor learning, 1047–1048
    perceptual learning, 1047–1048
    repetition priming, 1048
  motor control, 607–608
  regionalization maps, 34
Cerebral metabolic rate (CMR), 261
  developmental changes, 922
Cervicocervico reflex, 635
CGN, *see cis*-Golgi network
Chandelier cell, 46
Characteristic frequency (CF), auditory nerve, 562, 567
Chemoreceptors, 761–762
Cheyne–Stokes respiration, *753*, 762
Chick embryo, neuronal types in, *340*
Chlamydomonas reinhardtii, *213*
Cholecystokinin (CCK), 742
  food intake regulation, 779
  satiety signal, 770–772
Choline acetyltransferase (ChAT), 445
Cholinergic basal forebrain (CBF) neurons, *1032*, 1095
Chomsky's model, 1059
Chordamesoderm (CM), 293
Chordin gene, 289
Chorea, 664
Choroid plexus, function, 26, *28*, 36
Chromaffin cells, 343
Chronopharmacology, *838*
CICR, *see* Calcium-induced calcium release
Ciliary neurotrophic factor (CNTF), activities, 426
Cinematographic vision, 1098
Circadian misalignment disorders, 864–865
Circadian photoreception in mammals, 832–836
Circadian rhythm, 819
  chronopharmacology, *838*
  clock genes, 824–832
  Drosophila studies, *823*, *826–828*
  flight orientation in butterflies, *842*
  jet lag and shift work, *840*
  knockout mouse studies of redundancies, 825–831
  neurohormone release, 802–803
  peripheral oscillators, 824
  sleep disorders, *829–830*
  suprachiasmatic nucleus control
    entrainment, 832, 835–836
    humoral outputs, 837
    neuroanatomy, 819
    oscillator and pacemaker function, 821
    outputs
      amplification and divergence, 837–838
      autonomic function regulation, 840–841
      humoral outputs, 837
      neuroendocrine function regulation, 839–840
      pineal function regulation, 837
      sleep-wake cycle regulation, 838–839
    retina
      input to nucleus, 835
      photoreceptive system, 832–835
    rhythmic firing of neurons, 836–837
    single-cell oscillators, 821–824
  transcriptional feedback loop, 824–832
Circle of Willis, 36
Circulatory system, *see* Cardiovascular system
Circumventricular organs, 721–722
cis-Golgi network (CGN), 68
Classical conditioning, 1010, *1010*, *1015*
Clathrin-coated vesicles, 154
Climbing fiber, 452
  synapse, 685
Clock, circadian control, 825, *828*, 831
CM, *see* Chordamesoderm
CMR, *see* Cerebral metabolic rate
CNS, *see* Central nervous system
CNTF, *see* Ciliary neurotrophic factor
Coat proteins (COPs), protein trafficking, 68
Coated vesicles (CV), *252*
Cocaine, addiction mechanisms, 883–884
Cocaine-amphetamine-related transcript (CART), food intake regulation, 777
Cochlea
  central nucleus unit classification
    contralateral *versus* ipsilateral ear response, 568–569
    poststimulus time histogram, 566–567
    response map, 567–568
  endolymph, 559
  hair cells, 554–555
  implants, *561*
  outer hair cell mechanical properties and amplification, 557–559
  receptor potentials from inner hair cells, 556–557, *558*
  sensorineural hearing loss, 559
  stereocilia channels and transducer gating, 555–556
  structure, 554, *555*

Cognitive control, 1069, 1088
Cognitive development
  declarative memory, 924–925
  face processing, 925
  pathology, 934–940
Collateralization, nerve net, 17
Comatose, 1094
Commissureless, 379
Communication, *see also* Language
  animal lessons for human language, 1056–1057
  innate and learned aspects of animal, 1054–1055
  neural structures, 1056
  primates, 1055–1056
  strategies and mechanisms in animals, 1053–1054
  survival importance, 1053
Computational modeling, *232–233*
COMT, *see* Catechol-*O*-methyltransferase
Conditional specification, 407
Conditioned response (CR), 1010
Conditioned stimulus (CS), 1009, *1010*
Conditioning, 871–873
  eyeblink classical, 1043–1044
  fear, *1045*
Cone
  axon, *see* Growth cone guidance
  eye, 581, 585–586
Congenital central hypoventilation syndrome (CCHS), *764*
Congenital myasthenia, 385
Consciousness, 1091
  animal features, 1093
  arousal, 1093–1095
  content, 1093
  forward *versus* feedback projections, 1099–1100
  free will neurobiology, 1092
  information theory, 1100–1101
  neuronal correlates of consciousness, 1096–1097
  perceptual illusion, 1097–1098
  phenomena, 1091–1092
  states, 1093–1095
Consolidation, retrograde amnesia and, 1039–1040
Constructional apraxia, 959, 970–972, *972*, *978–979*
Convergent-extension (CE), neural induction, 295
COPs, *see* Coat proteins
Cortex, patterning, 305–306
Cortical motor areas, 641–642
Corticospinal neurons, 607
Cortico-striato-pallido-thalamic (CSPT), *673–674*
Corticotropin-releasing hormone (CRH), 839
  food intake regulation, 780
  neurochemical adaptation in addiction, 888–889
  release patterns, 803–804
CPEB protein, *see* Cytoplasmic polyadenylation element binding protein
CPG, *see* Central pattern generator
Cranial nerves, 34, *34*
Craniosacral division, *see* Parasympathetic nervous system
Craving, types, *891*
CREB, *see* cAMP responsive element binding protein
CRH, *see* Corticotropin-releasing hormone
Cryptochrome, circadian control, *827*, 831
Cuneate fasciculus, 543
Cut, dendrite development role in *Drosophila*, 459
CV, *see* Coated vesicles
CVLM, *see* Caudal ventrolateral medulla
Cycle, circadian control, *827*
Cyclic adenosine monophosphate (cAMP), 1011
Cyclic adenosine monophosphate responsive element binding (CREB) protein
  dendrite development role, 471
  long-term sensitization in *Aplysia*, 1012
Cyclic GMP phosphodiesterase, G protein-coupled receptor signaling, 196
Cystic fibrosis, ion channel mutation, 107
Cytoplasmic polyadenylation element binding (CPEB) protein, 472
Cytoplasmic proteins, 74–75
Cytoskeleton
  actin cytoskeleton, cell movement and transport, 78–80
  growth cone dynamics, 374–378
  and guidance receptors, 376–377
  interactions of systems, 82
  intermediate filaments, 80–82
  microtubule structure and function, 75–82
  microtubule-associated proteins, 76, *77*
  Rho and dendritic cytoskeleton structure control, *464–465*

D

DA, *see* Dopamine
DAG, *see* Diacylglycerol
DBH, *see* Dopamine-β-hydroxylase
DBL, *see* Dorsal blastopore lip cells
Declarative memory, 924–925, 986, 1029, 1031, 1033, 1035–1039, 1046–1048
Declarative spatial memory formation, medial temporal lobe, 984
Dedepression, synaptic strength, 1023
Deep brain stimulation (DBS), 667, *671–672*
Deep sequencing methods, 10
Dehydration anorexia, 772
Delayed sleep phase syndrome (DSPS), 829
Delta, neuronal specification role, 313, *313*, 322, *324*, 325–326
Dementia, 940–943
Dendrite, 4, 41
  active, synaptic integration
    dendritic spikes, 244–249
    inhibition and dendritic excitability, 249–250
    voltage-gated channels, 242–244
    weakly excitable dendrites, 240–242
  Boolean operations and dendritic arithmetic
    CA3 pyramidal neurons, *255*, 256
    dendritic spikes, *255*, 256
    distal apical dendrites, *255*
    neuromodulatory systems, 256
    NMDA, *255*, 256
    pyramidal cells, *255*
    synaptic integration, 254
  development
    activity-dependent, 469
    branches and synapse elimination, 471
    circuit formation in mammalian retina, 466
    convergence and divergence, 472–473
    extracellular regulation
      growth and branching regulation, 461–466
      orientation of dendrites, 460–461
    inner plexiform layer, 466–469
    protein synthesis dependent, 472
    transcriptional control in *Drosophila*, 458–460
  functional overview, *65*
  and neuronal response properties
    EPSPs, 257
    NMDA receptors, 257
    pyramidal neurons, 257
  passive, synaptic integration
    filtering, functional implications of, 239
    interactions between synapses, 237–239
    kinetics effects, 237
    potentials at soma, 234–237
  structure, 64, 231, *232*
Dendritic calcium signaling, *243*
Dendritic excitability, *248*, 248
  coincidence detection, 247
  intracellular recordings, 239
  pyramidal neurons, 240
  synaptic plasticity, 254
  voltage-gated channels, 242
Dendritic filtering
  EPSP and EPSC, 237, *238*
  functional implications of, 239
  membrane capacitance, 237
  voltage-clamp electrode, 237
Dendritic inhibition, *250*
  CA1 pyramidal neuron, 250, *250*
  and excitability, 249–250
  on-path, *238*

Dendritic recordings, *235*, 239, *240*
Dendritic spikes, 459
  Ca^{2+}, 244, *246*
  functions
    coincidence detection, 247
    Hebbian synaptic plasticity, 248
    multiple dendritic domains, 247
    synaptic integration, 247
    in vivo and *in vitro*, 247
  Na^{+}, 244–245, *245*
  NMDA, 244, *246*, 247
  propagation, 244
  types, 244, *245*
Dendritic spines, 43–45
  chemical and electrical compartmentalization, 251
  AMPA receptors, 251
  CA1 pyramidal neuron, *253*
  EPSP, *253*
  high-resistance spine, 254
  NMDA receptors, 251, 254
  voltage-gated channels, 254
  structure and function, 250–254
Dendro–dendritic interactions, 467–469
Dentate, cerebellar nucleus, 689–690
Dentate granule cells (DG), 400, *401*
Dependence receptors, 405
Depolarization, movements of ions, 95–97, 112
Depotentiation, synaptic strength, 1023
Depression, synaptic, 156, 158–160, *159–160*
D-hair receptor, 536
Diabetes insipidus (DI), *788*, *794*
Diacylglycerol (DAG)
  G protein-coupled receptor signaling, 189, 193
  schematic pathway of, *194*
Dishabituation, nonassociative learning, 1010
Disinhibition, 1072–1073
Diuresis, 790
Diurnal patterns, neurohormone release, 802–803
Dok-7 adapter protein, 385, 391, *392*
Dopamine (DA), *see also* Basal ganglia
  biosynthesis, 120, *121*, 122
  hypothalamic, 813
  mesolimbic dopamine system and addiction
    cocaine, 883
    neural substrates of sensitization, 884–885
    nicotine, 886
  neurochemical adaptation in addiction, 888–889
  orbitofrontal cortex, *656*, 669, *673–674*
  receptors, 184
  release
    autoreceptor regulation, 123–124
    enzymatic inactivation, 124
    neuronal transporters, 124–125
  schizophrenia defects, 938
  sleep control, 858
  SNpc input to striatum, 660
  storage
    vesicular monoamine transporters, 122–123
    vesicular storage, 122
  supplementary motor area, *656*, *675*
  transporter, 124–125
Dopamine-β-hydroxylase, catecholamines biosynthesis, 122
Dopaminergic neurons, 48–49, 607, 609, 876, 885
Dopaminergic syndrome of neglect, *876*
dOR, *see Drosophila* odorant receptors
Dorsal blastopore lip (DBL) cells, 287
Dorsal cortex of reptiles, *905*
Dorsal migrating cells, *344*
Dorsal organizers, DV axis, 303–304
Dorsal processing streams, 959, *960*
Dorsal respiratory group (DRG), 751
Dorsal root ganglion (DRG), 343
  neuron features, 64
Dorsoventral (DV) axis, 287
  dorsal organizers, 303–304
  homeobox genes, 304
  local organizers of, 302
  telencephalon, 305
  ventral organizers, 302–303
Double bouquet cell, *46*
Doubletime, circadian control, *828*
Dreaming, 859–861, *860*
  functions of, 866–867
DRG, *see* Dorsal respiratory group; Dorsal root ganglion
Drosophila, 391
Drosophila melanogaster, 368, 396
Drosophila odorant receptors (dOR), 525
Dscam, dendrite development role in *Drosophila*, 469
DSPS, *see* Delayed sleep phase syndrome
Dura, 35
Dynein, 83
Dynorphin-κ opioid systems, 888–889
Dysdiadochokinesia, definition, *688*
Dysexecutive syndrome, 1071–1072
Dyslexia, *935*
Dysmetria, definition, *688*
Dystrophin, functions, 79

E

Ear, *see* Audition
Early gastrula organizer (EGO), 292
Early-long-term potentiation (E-LTP), maintenance of, 1020–1021
EBN, *see* Excitatory burst neuron
ECF, *see* Extracellular fluid
Echolocation, and bats, *573–574*
ECM, *see* Extracellular matrix
Ectodermal cells, 287, *288*, 295
EEG, *see* Electroencephalography
EGF, *see* Epidermal growth factor
Electroencephalography (EEG), 847, *849*, 859
Electromyography (EMG), *627*, 644
Electro-olfactogram (EOG), 522
Electroplaque, *386*
Embryonic stem (ES) cells, 289
Emotional motor system, *745*
Empty spiracles genes, 298
End plate potential (EPP), 211, *618*
Endbulb, central auditory processing, *568*
Endocrine-disrupting chemicals (EDCs), *816*
Endolymph, cochlea, 559
Endothelial cells, 56
Energy metabolism, *see* Brain energy metabolism
Enteric nervous system (ENS), 736–738
EOG, *see* Electro-olfactogram
E.P., 984, *1031*, 1039
  memory impairment and brain damage, 1033
Eph receptors, *346*
EphrinBs, 399
Ephrins, growth cone guidance, 372
Epidermal growth factor (EGF), 389
Epinephrine
  biosynthesis, 120
  release
    autoreceptor regulation, 123–124
    enzymatic inactivation, 124
    neuronal transporters, 124–125
  storage
    vesicular monoamine transporters, 122–123
    vesicular storage, 122
  transporter, 124–125
Epithelial-to-mesenchymal conversion, 340
EPP, *see* End plate potential
EPSC, *see* Excitatory postsynaptic current
EPSP, *see* Excitatory postsynaptic potential
ES cells, *see* Embryonic stem cells
Estrogen-response elements (EREs), 811
Ethics, research, 11
Ets domain transcription factors, 394
Event-related potentials (ERPs), *993*
Evolution
  apes and hominins, *908*
  hominin brains, 914–916
  primates, 906–909, *907*
Excitation-secretion coupling, 142
  calcium binding and exocytosis trigger, 143–145
  calcium microdomains and vesicle release, 143
Excitatory burst neuron (EBN), 705–706, *706*
Excitatory postsynaptic current (EPSC), 222, *222*
Excitatory postsynaptic potential (EPSP), 139, 212, *214*, 1013
Extinction, attention deficit, 991
Extracellular fluid (ECF), 783–784
Extracellular matrix (ECM), 340, 351, 375
Extracellular positive feedback loop, 1013

Extrastriosomal matrix, 658
Eye, 579
  movement, 605
    central control, 701–703
    damage to SNpr causes involuntary, 665
    disorders, *708*
    fixation and fixational eye movements, 699
    gaze redirecting, 698–699, 705–707
    gaze stabilization, 697–698, 703–705
    muscle pairs, 700–701
    saccades and pursuit, 711
    shifts of visual attention, 711–712
    SNpr, *709*, 711
    substantia nigra output, 659
    visual perception and cognition contribution, 712–713
  optics, 579–580
Eyeblink classical conditioning, 1043–1044
Eyelid conditioning and cerebellum, 691–692

F

FA, *see* Friedrich's ataxia
Face processing, development, 925–926
Face recognition, prosopagnosia, 960–961
Facilitation, synaptic, 158, *159*, 160
F-actin, 374, 376
Fasciculation selective, 468–469
  organization of dendritic field, 468–469
Fastigial, 678, *682*, *688*, *691*, 693
Fastigius, cerebellar nucleus, 689
Fast-spiking interneurons (FSIs), 662, 664
Fear conditioning, *1045*
Feature integration theory (FIT), 1001
Feedback, *680*, 683–685, 689, 693, *694*
Feedforward, 683, 690, *690*, 693–694, *694*
Feed forward control, 602
FEF, *see* Frontal eye field
FFA, *see* Fusiform face area
Fibrillar actin, *see* F-actin
Fibroblast growth factor (FGF), 373, 396
  neural induction, 290–292
Filial imprinting, *492*
Filopodia, 363
Firing rate rhythms, 836–837
Flash suppression, 1097
Flat bipolars, 581
Flexor reflex afferent (FRA), 628
Fluid intake
  osmotic homeostasis regulation
    arginine vasopressin, 785, *789*
    diuresis, 790
    natriuresis, 789–790
    osmoreceptor cells, 787–789
    thirst, 785
  physiology of body fluids, 783–784
  rapid inhibitory feedback control, 786–787
  volume homeostasis regulation
    aldosterone secretion, 793–795
    arginine vasopressin secretion signals, 790–792
    hypovolemia signals, 792
    oxytocin, 795
    salt appetite, 792–795
Flutter, lower-frequency vibration, 534
fMRI, *see* Functional magnetic resonance imaging
Folium, cerebellum, 678
Follicle-stimulating hormone (FSH), 808
  release pattern, 811
Food intake
  body weight influences, 772–774
  brain stem regulation, 780–781
  glucostatic hypothesis, *773*
  hypothalamus regulation, 726, 776–779
  peptide regulation, 779–780
  satiety signals, 769–770
    cholecystokinin, 770–772
    experimental approaches, *771*
    gastric distension, 770–772
    postgastric effects of meals, 772
Forebrain, 293
c-Fos
  addiction role, 893
  gene activation, 207
FosB, addiction role, 893, *894*
Fossil, of brain, *902*
Fovea, 502
FRA, *see* Flexor reflex afferent
Fragile X syndrome, *936*
Free will neurobiology, consciousness, 1092
Friedrich's ataxia (FA), *937*
Frontal cortex, 34, 978–979
  fronto-parietal attention control system, 994–998
  value representation and motivational modulation, 982–984
Frontal eye field (FEF), *709*, 1088
  attention control system, 994–998
  electrical stimulation, *1000*
  gaze shift control, 710
Frontal lobe, subthalamic nucleus receives inputs, 658
FSH, *see* Follicle-stimulating hormone
Functional magnetic resonance imaging (fMRI), *266*, 959, *961–962*, *964*, 1035
  attention studies, 993, *993*, 1000, 1004
  consciousness, 1092, 1094, 1098, *1100*
  hippocampus memory system, 1035
  language studies, 1060
  prefrontal cortex, 1074–1075, 1081
Functional polarity theory, 17, *20*
Fusiform face area (FFA), 961, *962*, *964*
Fusion mechanism, 146
Fyn, dendrite orientation role, 461

G

G protein-coupled receptors (GPCRs), 163, *164*, 177, *see also specific receptors*
  desensitization, 180–181
    loss of surface receptors, 182
    phosphorylation, 181–182
  determinant for, 179
  glutamate receptor, 185–186
  homo and hetero oligomers, 180
  neurotransmitter-binding site, 177–179
  posttranslational modifications, 182
  receptor and activation of, 179
  signaling
    adenylate cyclase
      coincidence detection, 193
      fine-tuning of cAMP, 192
      inhibition, 192
      isoforms of, *192*
      receptor coupling, 192–193
      regulation of, 192–193
    calcium, 195
    conversion of, 189–191
    cyclic GMP phosphodiesterase, 196
    G protein cycle, 190
    G protein subunits, 190–191
    guanylate cyclase, 195–196
    ion channel modulation by G proteins, 196
    nitric oxide, 195–196
    response specificity, 192
  specificity and potency of, 180
  structure, 177
GABA, *see* γ-Aminobutyric acid
$GABA_A$ receptor, 128
  binds drugs, 170
  glycine receptor, 170–171
    cluster at synapses, 171
  inhibitory function, 169–170
$GABA_B$ receptor, 128, 186
Galago garnetti, *912*
Gambling task, *1073*
Gamma-band oscillations, 1038
Gamma-band spike-field coherence, in monkey hippocampus, *1038*
Gamma-band synchronization, 1038, *1038*
Ganglia, *344*, 451, *452*
Ganglion mother cell (GMC), *310*, *317*
  neurogenesis, 312
  Notch pathway activity, 315
Gastrointestinal phase, 768, *769*
Gaze, *see* Eye
GDNF, *see* Glial cell-derived neurotrophic factor
Gene expression, electrical activity regulates, 394
Generator potential, *532*
Genetic diseases affecting breathing, *764*
Genomics, 9–10
Gephyrin, 398
GF, *see* Glial fiber
GFAP, *see* Glial fibrillary acidic protein
GFR, *see* Glomerular filtration rate
GH, *see* Growth hormone
GHRH, *see* Growth hormone-releasing hormone
Glia, 276–277
  cells, 136

neuron support in brain, 7
neurons and, 4–6
Glial cell-derived neurotrophic factor (GDNF), 373
activities, 426
Glial fiber (GF), *353*
cellular organization of neurons, *353*
Glial fibrillary acidic protein (GFAP), astrocyte function, 51
Glial-guided neuronal migration, *353*
Gliotransmission, 136–138
Globus pallidus, *see* Basal ganglia
Globus pallidus external segment (GPe), 660, 663
Globus pallidus internal segment (GPi), 659–660, *659*, 662–663
Glomerular filtration rate (GFR), blood pressure response, 784
Glucagon-like peptide-1 (GLP-1), *771*
Glucocorticoid receptors (GR), 802, 804
temperament shaping in rat, 489–491
Gluconeogenesis, 768
Glucose, 767, *768*, 772
transporters, brain energy metabolism, 277–278
Glucose metabolism, *see* Brain energy metabolism
Glucostatic hypothesis, *773*
GLUT-1, blood-brain barrier function, 57
Glutamate
astrocyte metabolism, 274–276
excitatory neurotransmission, 128–129
metabolism, 274–276
Glutamate receptors, 171, 173, *see also specific receptors*
cluster at synapses, 175–176
GPCRs, 185–186
non-NMDA receptors, 171–173
RNA splicing and editing, 173
Glutamine synthetase, astrocyte function, 51
Glutathione, reactive oxygen species protection, 270–271
Glycine receptor, structure, 170–171
Glycogen, 768
astrocyte metabolism, 282
synthase kinase 3 (GSK3), 291
Glycogenolysis, 768
Glycolysis, brain energy metabolism, 269
GMC, *see* Ganglion mother cell
Goldman–Hodgkin–Katz equation, *98*
Golgi apparatus, 66–67, *353*
protein trafficking, 68
trans-Golgi network
endocytosis and membrane cycling, 71–72
protein exit, 70–71
Golgi cell, *681–682*, 682–685, *693*
Golgi staining, 31
Golgi tendon organ, proprioception, 536, 620
Gonadotropin-releasing hormone (GnRH), 724, 808, *809*
receptor, 812
release patterns, 811
GPCRs, *see* G protein-coupled receptors
GR, *see* Glucocorticoid receptor
Gracile fasciculus, 543
Gray matter, spinal cord, 33
Green fluorescent protein (GFP), 341
Grey rami communicans, 731
Grid cells, 985, *985*, 1036, *1037*
Growth cone guidance, 363–365
axon trajectory segments, 366–374
cues
ephrin, 372
morphogens and growth factors, 372–373
netrin, 368–370
semaphorin, 370–372
slit, 372
cytoskeletal dynamics, 374–378
midline of developing nervous systems initial attraction, 378–379
premature repulsion prevention, 379–380
switching, 379
study approaches, 366
Growth hormone (GH), 806, *807*
receptor, 808
Growth hormone-releasing hormone (GHRH), 806
GSK3, *see* Glycogen synthase kinase 3
Guanylate cyclase, G protein-coupled receptor signaling, 195–196

H

Habenula, 876, 887
Habits, memory for, 1042–1043
Habituation, 1010
Hair cells
damaging, 559
descending systems to, 564–565
increasing sensitivity of hearing, 559
mechanical energy translation into electrical energy, 554–555
Hair follicle afferent (HFA), 536
HapMap project, 10
HD, *see* Huntington's disease
Head-direction cells, 720, 985
Hearing, *see* Audition
Hebbian
mechanism, 1018
synaptic plasticity, 248, *248*
Hebb's postulate, synapse elimination, 447
Hedgehog (Hh), growth cone guidance, 372
Hemiballismus, 664
Hemispatial neglect, 972–973, 990–992
Hemisphere, and language processing, 1061–1062
Hepatic encephalopathy, astrocyte dysfunction, *275–276*
Hepatocyte growth factor (HGF), growth cone guidance, 373
Heteroceptor, 119
Heterosynaptic facilitation, 229, 1011, *1011*, 1016
HFA, *see* Hair follicle afferent
HGPPS, *see* Horizontal Gaze Palsy with Progressive Scoliosis
Hh, *see* Hedgehog
Hierarchy
motor system, 27–29
nervous system operations, 5–6
High-frequency stimulation (HFS), 1016
Hindbrain segmentation, 296–297
Hippocampal CA1 pyramidal neuron, *252*
Hippocampus, 231
grid cells, 985
head-direction cells, 985
and memory system
amnesia, 1031–1033
anatomy, 1030–1031, *1031*
anterograde amnesia animal models, 1033–1034, 1039
declarative memory, 1035–1039
long-term potentiation and CA3-CA1 synapse, 1016–1018
neuroimaging, 1035
recognition memory signals, *1037*
retrograde amnesia and consolidation, 1039–1040
transgenic animals and dissection, 1039
place cells, 985
postnatal refinement of synaptic contacts, 921
spatial cognition role
declarative spatial memory, 984
memory formation, 984
neuron place fields, 984–985
parahippocampal and retrosplenial cortex, 984
Histamine, and sleep control, 858
H.M., 984
and amnesia, 1031–1033
Hominins, 908
Homo florensensis, 908
Homo habilis, 908
Homo sapiens, 908, 914
Horizontal cell, 582–583
Horizontal Gaze Palsy with Progressive Scoliosis (HGPPS), 379
Hormonal regulation, 483, 815–816, 865
Horse-radish peroxidase (HRP), 146
Hox genes, cellular determination, *311*, 328, 330, *331*
H-reflex, *627*
Human brain, basal ganglia in, *654*
Human lifespan, changes in sleep over, 861–862
Humans, pheromone detection in, 528
Huntington's disease (HD), pathophysiology of, 664, *671*

Hydra, *see* Nerve net
6-Hydroxydopamine (6-OHDA), 665
Hyperphagia, 773
Hyperpolarization, 95–97, 114–116
Hyperpolarization-activated cation (HCN) channels, 242
Hypothalamic hunger, 776–777
Hypothalamic-lactational axis, *812*, 812–813
  disruptions in, 813
Hypothalamic-pituitary-adrenal (HPA) axis, *803*, 803–804
  disruptions in, 804
Hypothalamic-pituitary-gonadal (HPG) axis, 808–811, *809*
  disruptions in, 812
  receptors, 811–812
Hypothalamic-pituitary-thyroid (HTP) axis, 804–806
  disruptions in, 806
  receptors, 806
Hypothalamus, 717–718
  appetite regulation, 726, 776–777
  autonomic function central integration, 724
  behavioral state control, 722–723
  consummatory behavior role, 873
  cytoarchitecture, 718–719
  fluid homeostasis and thirst, 725–726
  functional organization
    circumventricular organs, 721–722
    integration of physiology and behavior, 720–721
    limbic projections, 719
    olfaction, 719
    visceral sensation, 720
    visual projections, 720
  historical perspective on, *800*
  immune function, 724–725
  lateral hypothalamic syndrome, 874
  motivation role, 873
  motor control, 606
  neuroendocrine system, 799–800
    hormone release cycles, 802–803
    lactational axis, *812*, 812–813
    pituitary-adrenal axis, 803–804
    pituitary-gonadal axis, 808–811, *809*
      disruptions, 812
      gonadotropin-releasing hormone, 808
      receptors, 811–812
    pituitary targets, 800
    pituitary-thyroid axis, 804–806
      disruptions, 806
    posterior pituitary hormones, 813–815
    regulation by neural, glial, and hormonal inputs, 802
    somatotropic axis, 806–807
      disruptions, 808
      growth hormone-releasing hormone, 807–808
    pituitary gland neuroendocrine regulation, 723–724
    reproductive function, 724
    thermoregulation, 724–725
    ventrolateral preoptic area of sleep-wake cycles, 852
Hypothesis, creation, 11
Hypothyroidism, 806
Hypotonia, definition, *688*
Hypovolemia
  aldosterone secretion, 793–795
  neural and endocrine signals
    AVP secretion, 790–793
    thirst, 792–793
  salt appetite, 793–795

I

Ia afferents, 536
Ib afferents, 620
IBN, *see* Inhibitory burst neuron
ICF, *see* Intracellular fluid
IEGs, *see* Immediate early genes
IFG, *see* Inferior frontal gyrus
IGF-I, *see* Insulin-like growth factor-I
ILD, *see* Interaural level difference
ILN, *see* Intralaminar nuclei of the thalamus
Imaging dendritic function
  conventional electrophysiological techniques, *242*
  microelectrodes, *242*
  optical techniques, *243*
  two-photon excitation, *243*
  two-photon microscopy, *242*
IMM, *see* Intermediate and medial mesopallium
Immediate early genes (IEGs), transcription regulation, 207
Impulsivity, *883*
In vivo imaging, *243*
Induced pluripotent stem cells (iPSCs), *318*
Induction, LTP, 1018–1022
Inferior colliculus, central auditory processing, 570–572
Inferior frontal gyrus (IFG), 1062
Inferior mesenteric ganglion, 733
Inferior temporal cortex, agnosia
  early studies, 948
  object recognition, 950
Inferior temporal (IT) neurons, *1005*
Information theory, consciousness modeling, 1100–1101
Infundibulum, 23–24, *24*
Inhibitory burst neuron (IBN), 705–706, *706*
Inhibitory interneuron
  cerebral cortex, 45–47
  compilation of, *250*
Inhibitory postsynaptic potential (IPSP), 139, *214*, 218, 220–222
Inner ears, 553, *554*
Inner plexiform layer (IPL), 466–469
  organization of dendritic fields, 467–469
Inositol 1,4,5-trisphosphate (IP_3)
  G protein-coupled receptor signaling, 189
  schematic pathway of, *194*
  termination of, 194–195
Instrumental conditioning, 1009, 1077
Insulin
  adiposity signaling, 774
  caloric homeostasis, 768–769
Insulin-like growth factor-I (IGF-I), receptor, 808
Interaural level difference (ILD), 484, *569*, 569–570
Interaural time difference (ITD), 484–486, *569*, 569–570
Intergeniculate leaflet (IGL), 833
Intermediate and medial mesopallium (IMM), filial imprinting, *492*
Intermediate filaments, *80*, 80–82
Intermediate-phase memory, for sensitization, 1014
Intermediate-term memory, 1014
Intermediolateral nucleus, 731
Internal carotid artery, 36
Internal jugular vein, 37
Internal medullary lamina, 659
Interneurons (IN), 211, 1011
  cortical interneuron migration, 356
  flatworms, 19–20
  motor control, *see* Spinal intereurons
α-Internexin, 81
Internode, 49
Interpositus, cerebellar nucleus, 689
Intracellular fluid (ICF), 783–784
Intracellular neurofibrillary tangles (NFTs), *1032*
Intracellular positive feedback loop, 1013
Intrafusal fibers, 620
Intralaminar nuclei of the thalamus (ILN), consciousness, *1095*, 1095
Invaginating bipolars, 581
Ion channels, 107–108, *see also specific channels*
  diseases, *107–108*
  G protein modulation, 196
Ionotropic receptors, 163, *164*, *see also specific receptors*
  stretch reflex studies of ionotropic postsynaptic potentials, 211–214
IP_3, *see* Inositol 1,4,5-trisphosphate
IPL, *see* Inner plexiform layer
iPSCs, *see* Induced pluripotent stem cells
IPSP, *see* Inhibitory postsynaptic potential
ITD, *see* Interaural time difference

J

c-Jun, regulation, 207
Junctions, definition, 4
Just noticeable difference (JND), 500

K

Kainate receptor, 173
Katz model, 157–158
Kennedy's disease, *622*
Ketogenesis, 768
Kinesin, 83
Koniocellular layer, lateral geniculate nucleus, 584
Kugelberg-Welander disease, *622*

L

Labeled line principle
  sensory systems, 503–504, *504*
  taste system, 519
Labyrinth, vestibular apparatus, 632–633
Lactate
  astrocyte release, 281–282
  brain utilization, 263
Lactation, hypothalamic control of, 812–813
Lamellipodia, 363
Language
  acquisition and sensitive periods, *480*
  animal communication lessons, 1056–1057
  brain structures related to, 1060–1061
  development, 926–927
  evolution, 1059–1060
  lateralization, 1061–1062
  lexical access and word meaning, 1062–1064
  localizationist theories, 1059–1060
  perisylvian cortex organization for processing, 1062
  processing, 1057–1059
  syntactic operations, 1064–1066
Large aspiny neuron, striatum, 655
Larval salamander, amacrine cell, *232*
Late gastrula organizer (LGO), 292
Late-long-term potentiation (L-LTP), 1020–1021
  maintenance of, 1022
Lateral cervical nucleus, 544
Lateral ganglionic eminence (LGE), 305
Lateral geniculate nucleus (LGN), 909
  attention and neural response modulation, 1000–1001
  color-selective response in P pathway, 585–586
  experience effects on functional organization of visual cortex, 487
    critical period, 487–489
    plasticity mechanisms, 488–489
  input, 587
  layers and retinal ganglion cell projections in primates, 584
  M cells and contrast sensitivity, 586
  response properties of P an M pathways, 584–585
  retinogeniculocortical pathway
    extrastriate visual areas, 592–593
    lateral geniculate nucleus structure, 586–587
    parietal cortex, 593–594
    primary visual cortex
      cat, 587–589
      macaque, 589
    visual cortex
      columnar structure, 589–590
      cytochrome oxidase staining, 592
      optical imaging studies, 590–592
  role of vision, 577, 583, 586–587
Lateral hypothalamic syndrome cell body lesions, 874
Lateral inhibition, 578
Lateral intraparietal area (LIP), *709*, 710, 1083, 1088
  attention control system, 994
  neurons, *995*
  parietal cortex, 975–976
  reward probability, *997*
  spatial cognition, 975–976
  spatial salience representation, *995–996*
Lateral occipital complex (LOC)
  object recognition, 959
  shape processing, *961*
Lateralization, 1061–1062
LC, *see* Locus coeruleus
LDP, *see* Locomotor-drive-potentials
Learning and memory, 1029–1048, *see also* Long-term depression, Long-term potentiation
  aging effects on memory
    animal models, 928–929
    neurobiology of decline
      hippocampus, 929
      neuron loss, 929–930
      prefrontal cortex, 931–933
      subcortical systems, 933
    variability, 927–929
  Alzheimer's disease and, *1032–1033*
  amygdala memory system
    central nucleus, 1044
    emotional disposition acquisition, 1046
    fear conditioning, 1044, *1045*
    memory modulation, 1046–1047
  Aplysia studies, 1010
    associative learning of withdrawal reflexes, 1014–1016
    long-term sensitization, 1012–1014
    short-term sensitization, 1011–1012
    temporal domains for memory of sensitization, 1014
  associative learning, 1009–1010
  behavioral support by multiple memory systems, 1048–1049
  cerebellum memory system
    anatomy, 1043
    eyeblink classical conditioning, 1043–1044
  cerebral cortex memory system
    anatomy, 1047
    motor learning, 1047–1048
    perceptual learning, 1047–1048
    repetition priming, 1048
  conditioning, 871–873
  declarative memory development, 924–925
  developmental learning sensitive periods, 491
    closing, 493
    filial imprinting, *492*
    pathway maturation, 492–493
    prolongation via deprivation, 493
    reopening, 493–494
  hippocampus memory system, 1040
    amnesia, 1031–1033
    anatomy, 1030–1031, *1031*
    anterograde amnesia animal models, 1033–1035
    declarative memory, 1035–1039
    retrograde amnesia and consolidation, 1039–1040
    transgenic animals and dissection, 1039
  motor learning, 611
  nonassociative learning, 1010
  patient H.M., and amnesia, 1031–1033
  place *versus* response, *1042*
  striatum memory system, 1043
    actions, outcomes and habits, 1042–1043
    anatomy, 1041
    memory identification, 1041–1042
  synaptic strength and complex memory storage, 1023–1025
Leptin, adiposity signaling, 774
Leucine-rich repeat protein 2 (LRRTM2), 397
Leukemia inhibitory factor (LIF) activities, 426
LGE, *see* Lateral ganglionic eminence
LGN, *see* Lateral geniculate nucleus
LGO, *see* Late gastrula organizer
LH, *see* Luteinizing hormone
Limbic system
  hypothalamic projections, 719
  prefrontal cortex connectivity, 1070
Linguistic universals, 1059
LIP, *see* Lateral intraparietal area
Lipogenesis, 768
Lipolysis, 768
Lissauer's tract, 545
LMAN, *see* Anterior nidopallium
LOC, *see* Lateral occipital complex
Local circuit neurons, 6
Local interneuron, function, *32*
Locomotion, *see* Motor system
Locomotor activity circadian rhythms, 837
Locomotor-drive-potentials (LDP), *625*
Locus coeruleus (LC), 6, 547, 851, *853*, *857*, *875*, *884*, *1032*, 1047
Locust, mesothoracic ganglion of, *232*
Long hierarchical neurons, 6

Long QT syndrome, ion channel mutation, *108*
Long-term depression (LTD), 158, 929, 1023
  purkinje cell, synapse, 686–687
Long-term facilitation (LTF), *759*, 762, 1012–1013, 1020
Long-term potentiation (LTP), 158, 422, 929
  expression, 1021
  hippocampal CA3-CA1 synapse, 1016–1018, *1017*
  induction
    calcium, 1018–1019
    late potentiation, 1020–1021
    N-methyl-*D*-aspartate receptor, 1019–1020
  learning linkage, 1022–1023
  and LTD, 1023
  maintenance, 1022
Long-term sensitization, 1012–1014
Low-density lipoprotein receptor-related protein 4 (Lrp4), 385, *388*, 391
Lower-frequency vibration, 534
Lrp4, *see* Low density lipoprotein receptor-related protein 4
LRRTM2, *see* Leucine-rich repeat protein 2
LTD, *see* Long-term depression
LTP, *see* Long-term potentiation
Lugaro cell, 683
Luteinizing hormone (LH), 808
  receptors, 813
  release pattern of, 811
Lysosome, affect nervous system, 72

M

mAChR, *see* Muscarinic acetylcholine receptor
Macrocircuitry, nervous system, 31
Maculae, 632
Magnetic resonance imaging (MRI), 268, *895*, 1040
Magnetic resonance spectroscopy (MRS), brain metabolism studies, 263
Magnification factor, 508
Magnocellular layer, lateral geniculate nucleus, 584, 586
Major depressive disorder (MDD), *894–895*
Mammalian nervous system, synapse elimination in, *439*
Mammalian outer hair cell, mechanical properties of, 557–559, *558*
MANF, *see* Mesencephalic astrocyte-derived neurotrophic factor
Mantle layer, formation, *26*, 26–27
MAO, *see* Monoamine oxidase
MAPK, *see* Mitogen-activated protein kinase
MAPs, *see* Microtubule-associated proteins
Masking in auditory nerve response, 564
Maturity onset diabetes of the young (MODY), ion channel mutation, *107*
MBP, *see* Myelin basic protein
MCH, *see* Melanin-concentrating hormone
MCTs, *see* Monocaroxylate transporters
MDD, *see* Major depressive disorder
Medial ganglionic eminence (MGE), 305
Medial geniculate, central auditory processing, 572
Medial longitudinal fasciculus, 701
Medial superior olive (MSO), central auditory processing, 570
Medial superior temporal area (MST), pursuit eye movement role, 711
Medial temporal lobe
  declarative spatial memory, 984
  neuron place fields, 984–985
Median eminence (ME), *718*, 721, *721–722*, 723, *725*
Mediolateral axis, *21*
Medium spiny cell, 47
Medium spiny neuron, striatum, 655, *655*, 658, *658*
Meissner corpuscle, 534–535
Melanin-concentrating hormone (MCH), food intake regulation, 779
Melanocyte, 343
α-Melanocyte-stimulating hormone (α-MSH), food intake regulation, *777*
Melanopsin, 832, *834*
Melatonin
  receptor, 843
  release and function, 841
Membrane potential
  generation, 94
  hyperpolarization *versus* depolarization, 95–97
  ion pump maintenance of gradients, 97–98
  passive distribution of ions, 94–95
  resting membrane potential
    determinants, 97
    neuron-type dependence, 97
Membrane proteins, classification, 66
Membrane trafficking
  ATPase, 152
  membrane-bound compartments, 146
  neurons, 145–146
  SNARE, 150
Membranes, pre- and postsynaptic, 386, *386–387*
Memory, *see also* Learning and memory
  declarative, 924–925
  factors in long-term, 1012
  formation, 984
  interdependence, 865–866, *865*
  song learning and, 483
  striatum, *see* Striatum, memory system
  temporal domains for, 1014
Meniere syndrome, *638*
MEP, *see* Motor evoked potential
MEPSP, *see* Miniature excitatory postsynaptic potential
Merkel's disk, 535
Mesencephalic astrocyte-derived neurotrophic factor (MANF), 426
Mesencephalic locomotor region (MLR), 628
Mesolimbic dopamine system, 877–879
Mesolimbic system, 938
Metabolic fuel, 767–768
Metabolism, *see* Brain energy metabolism; aloric homeostasis
Metabotropic receptors, 119, *see also specific receptors*
  slow synaptic potential mediation, 224–226
1-Methyl-4-phenyl-1,2,5,6-tetrahydropyridine (MPTP), *667*
MGE, *see* Medial ganglionic eminence
Microcircuitry, nervous system, 31
Microglia, 54–55
  immune response mediation in nervous tissue, 54
  reactive microglia, 55
Microtubule, structure and function, 75–82
Microtubule-associated proteins (MAPs), 76, *77*
Midbrain, 296
Midbrain–hindbrain boundary (MHB), 299
Middle cervical ganglion, 732
Middle temporal area (MT), pursuit eye movement role, 711
Migration
  central nervous system cells
    cortical histogenesis, 348–350
    neurogenesis, 351–354
  embryonic zone formation
    molecular mechanisms of cell motility, 351–354
    primary proliferative matrix cell movements, 351
  peripheral nervous system cells, *see* Peripheral nervous system
  radial glial fibers, neuronal
    cerebellar system, 357
    cortical interneurons, 356
    layer 1 cells, *349*
    lineage analysis, 355–356
    RMS, 356–357
Mineralocorticoid receptor (MR), 804
Miniature excitatory postsynaptic potential (MEPSP), 139
Miniature inhibitory postsynaptic potential (MIPSP), 139
Minimal conscious state (MCS), 1094
Mirror-neurons, 607
Mitochondria, mechanisms for, 73–74
Mitogen-activated protein kinase (MAPK), 291, 1011–1012
  dendrite development role, 471
  neurotrophin receptor signaling, 424

MLR, *see* Mesencephalic locomotor region
Model
  computational, 677, *682*, 690–691, *691*, 693
  forward, *688*
MODY, *see* Maturity onset diabetes of the young
Molecular feedback loop, 836–837
Molecular genetic techniques, 254
Molecular mechanism for circadian rhythms, 824–832, *826–828*
Molecular motors in nervous system, 82–85
Monkey hippocampus
  gamma-band spike-field coherence in, *1038*
  recognition memory signals in, *1037*
Monoamine oxidase (MAO), 124
  inhibitors, *125*
Monocaroxylate transporters (MCTs), 263
Monosynaptic stretch reflex, 211
Morphogenetic furrow (MF), 324
Morris water maze task, 1033, *1035*
Mosaicism, 468
Mossy fiber
  and climbing fibers, 687–689
  glomerulus, 685
Motivated behavior, 879
Motivation, 871–879
  amygdala role, 879
  electrical stimulation of brain, 876–877
  feeding, *872*
  hypothalamus role
    behavior and dopaminergic axons, 873–876
    consummatory behavior, 873
    motivation, 873
  learning control of motivated behavior, 879, *880*
  mesolimbic dopamine system, 877–879
Motoneurons, 600–601
Motor axon, 387
Motor control, basal ganglia for, 668–669
Motor cortex
  cortical motor areas, 641–642
  corticospinal projection, 638–640
  movement control
    cues, 647
    M1, 646–647
  somatotopic organization, 642–645
  spatial cognition role, 979–980
Motor evoked potential (MEP), *627*
Motor learning, 610–611, 686, 1047–1048
Motor neuron (MN), 640, 1011
  cellular determination, 330–332, *331*, *333*
  nerve net, 17–18, *17*, 19
Motor neuroprosthetics, *648*
Motor pattern generator (MPG), 29, *29*
Motor pattern initiator (MPI), 29, *29*
Motor system
  brain stem-spinal cord, 639
  cerebellum, 609–610
  descending control, 631, *see also* Motor cortex
    basal ganglia, 607–608, 653–676
    brain stem-spinal cord, 606, 628
    cerebral cortex, 607–608
    hypothalamus, 606
    lateral voluntary system, 638–650
    medial postural system, 631–638
    stem control, 628
  development, 599–600
  eye movements, 605
  feedback and feedforward control, 602
  fiber types, 616–617
  learning, 611
  locomotion cycle, 613–615
  motor unit, 617–620
  muscle
    afferents, 620–621
    contraction, 616
    stretch reflexes, 621
  neural control of human walking, *627–628*
  neurons, 600–601, 615–616, 625
  patterns, 603
  posture, 605, 631–638
  rhythmic movements, 605
  sensory modulation
    cutaneous reflexes, 629
    flexor reflexes, 628–629
  sensory receptors in movement control, 600–601
  spinal interneurons, 622
    central pattern generator network models, 622–624
    excitatory interneurons, 624–625
    left–right coordination, 626
    reciprocal inhibition between flexor and extensor muscles, 625–626
    Renshaw inhibition, 626
  visuomotor coordination and reaction time, 608–611
MPG, *see* Motor pattern generator
MPI, *see* Motor pattern initiator
MRI, *see* Magnetic resonance imaging
MRS, *see* Magnetic resonance spectroscopy
α-MSH, *see* α-Melanocyte-stimulating hormone
MSO, *see* Medial superior olive
MST, *see* Medial superior temporal area
MT, *see* Middle temporal area
Multiple cellular processes, 1010–1014
Multiple system atrophy, *744*
Muscarinic acetylcholine receptor (mAChR), 182–183
Muscle
  afferents, 620–621
  contraction, 616
  differentiation, 390
  eye, 700–701
  fiber types, 616–617
  prepatterning, *388*
  species diversity in neuromuscular systems, *614*
  stretch reflexes, 621
Muscle spindle, proprioception, 536
Mushroom shaped dendritic spines, 251
MuSK, 385, 390
  downstream of, 391
  rapsyn role, 392–393
  tyrosine phosphorylation of, 391
Mutant mice
  agrin, 390, 393
  lrp4, 391
  MuSK, 390–391, *390*
  rapsyn, 392
Myasthenia, *618*
Myasthenic disorders, *618*
Myelin, 49–51
  action potential propagation speed effects, 106
  proteins, 51
  synthesis, 49–51
Myelin basic protein (MBP), trafficking, 73
Myofiber, *389*
  synaptic nuclei of, 393–394
Myogenin, 394
Myosin, 82–85
  brain functions, 85
  types of, 82
Myotonic dystrophy (DM), *936*
Myxedema, 806

N

nAChR, *see* Nicotinic acetylcholine receptor
NADPH, *see* Nicotinamide adenine dinucleotide phosphate
NANC neurons, *see* Nonadrenergic noncholinergic neurons
Natriuresis, 789–790
NCAM, *see* Neural cell adhesion molecule
NCC, *see* Neuronal correlates of consciousness
NE, *see* Norepinephrine
Near response, 699
Necker cube, 1097
Necrosis, features, 412
Neglect syndrome, 990–992
Neocortex, 231
  origins, 901, *902*, 903–904, *904*, 906
Neocortical pyramidal neuron, *252*
Nernst equation, *96*
Nerve fibers, descending systems to, 564–565
Nerve growth factor (NGF), 416–417
  assays, 416
  history of study, 406, 416–418
  neurotrophic hypothesis, 406, 417
  receptor, 419–421
  retrograde transport, 423
Nerve net, 15–16
  centralization, 18

Nerve net (*Cont.*)
motor neurons, 17–19, *17*
sensory neurons, 16–17, *17*
Nerve terminal, 385–386, *387*, *389*, *392*
Nervous system, 3
building and maintaining cells of, 85–91
lysosome affect in, 72
molecular motors in, 82–85
Nervous tissue, protein synthesis in, 65–75
N-Ethylmaleimide-sensitive factor (NSF), 69
membrane fusion, 152
protein trafficking, 69
Netrin, growth cone guidance, 368–370
Neural activity, declarative memory, 1035–1039
Neural cell adhesion molecule (NCAM)
growth cone guidance, 373
long-term sensitization in *Aplysia*, 1013
Neural circuits, cerebellum
dentate, 689–690
fastigius, 689
interpositus, 689
vestibular nucleus, 689
Neural control, of sleep–wake cycles, 850–859
Neural crest cells, 287, 320, *321*
avian embryos, *344*
dorsal migrating cells, *344*
graft studies, 344–345
labeling, *342*
migration cessation, 346–347
migration initiation, 340
neural axis levels, *343*
origins, 339
regionalization along body axis, 340–343
segmental migration, 345–347
techniques, 340–343
Neural groove, 23, *24*
Neural induction
CE, 295
default model for, 288–290
embryonic origins, 287
signaling pathways, 290–292
Spemann's organizer, 287–288
Neural morphogenesis, 295–296
Neural organization, for language, 1060–1066
Neural patterning
AP patterning
homeobox gene expression, 298–299
local organizers of, 300–301
molecular basis, 293–294
spinal nerves, 301–302
classical models for, 292–293
cortex, 305–306
DV patterning, 294–295
dorsal organizers, 303–304
homeobox genes, 304
local organizers of, 302
telencephalon, 305
ventral organizers, 302–303
hindbrain segmentation, 296–297
intersection of AP and DV patterning mechanisms, 304–305
Neural plate, 23–24
Neural stem cell (NSC), properties of, *318–319*
Neural structures, for animal communication system, 1056
Neural synapses, LTP occurs in, 1016–1018
Neural tube
development, 23, 309–310, *310–311*, 313
regionalization, 24–26
Neural tube defect (NTD), 295
Neuroblasts, 309, *310–311*, 312
Neurodegeneration in AD, *1032*
Neurodevelopmental disorders, synaptic targets in, *398*
Neuroendocrine functions regulation, 839–840
Neurofilament (NF), *390*
Neurogenesis, 310–312
Neuroglia, definition, 49–55
Neuroligin-neurexin interactions, 397
Neuromelanin, 660
Neuromodulation, 851
Neuromuscular junction, and synapse elimination, *438*, 439, *440*
interaxonal competition and activity, 446–453
structural analysis, 443–446
Neuromuscular synapse
development of, 385–394
formation, signaling mechanisms for, *392*
postsynaptic membranes and substructural organization, 386, *386–387*
substructural organization of, 385–386
Neuromuscular system, 446–448
Neuron doctrine, *16–17*, *20*, *32*
Neuronal cable theory, 231, *232–233*
Neuronal cell adhesion molecule (NCAM), 1013
Neuronal communication, modes of, 117–118
Neuronal control, of *Drosophila* circadian behavior, *823*
Neuronal correlates of consciousness (NCC), 1096–1097
Neuronal migration, *349*
glial-guided, *353*, 355
non-radial pathways, *356*
radial glial cells, *350*, 354–357
Neuronal receptive fields, 1002–1004
Neurons
actions of NO on, 135–136
active conductances, 111
astrocyte-neuron metabolic unit, 283–284
brain organization and function, 6
classification, 4
electrophysiological diversity, 108–111
encode attributes, 974
energy-consuming processes, 271–273
functional categories, 4–5
in hippocampal role, 984–985, *985*
interneuronal communication, 4
morphology, 41–43
neuromodulatory, 850–851
operative processes, 5–6
passive electrical properties, 99–102
structural categories, 4
Neuropeptide Y (NPY)
appetite regulation, 726
food intake regulation, 778
Neuropil, definition of, 19, *19*, *32*, 33
Neuropsychiatric diseases, *829–830*
Neurotensin (NT)
coexistence with classical neurotransmitters, 132
inactivation, 132
synthesis, 132
Neurotransmitters, 118, 394 , *see also specific neurotransmitters*
in behavioral state control, 858–859
classical, 120–129
colocalization, 129–130
convergence of different transmitter-containing axons on neuron, 129
excitation-secretion coupling, 142–145
fast *versus* slow responses of target neurons, 130
history of study, 117
NO as, 136
nonclassical, 129–130
peptide neurotransmitters
inactivation, 130–132
synthesis and storage, 130
quantal analysis, *see* Quantal analysis
quantal release, 139–142
release mechanisms
calcium-binding proteins and action potential coupling to exocytosis, 153
endocytosis and recovery of vesicle components, 154–155
fusion rate, 146
membrane trafficking, 145–146
NSF, 152
similarity with other cell processes, 146–147
SNARE complex, 149–152
synaptic proteins, 147–149
transmitter packaging in vesicles, 153–154
vesicle docking and priming, 152–153
short-term synaptic plasticity, 158–161, *159*
steps in neurotransmission
receptor binding and recognition, 119
storage, 118
synthesis, 118
termination of action, 119

synaptic specializations *versus* neurons, 130
unconventional transmitters, 132–136
Neurotrophic factors (NTFs), 405
Neurotrophic hypothesis, 406, 417
Neurotrophins, 419
activities, 426–428
dendrite growth and branching regulation, 461–462
endosome signaling hypothesis of retrograde axonal signaling, 423
processing and secretion, 421–422
receptors, 419–421
signaling, 423–425
synaptic plasticity role, 422–423
Neurulation, 23, 287
NF, *see* Neurofilament
NGF, *see* Nerve growth factor
Nicotinamide adenine dinucleotide phosphate (NADPH), 269
Nicotine, addiction mechanisms, 886
Nicotinic acetylcholine receptor (nAChR)
assembly, 167–168
clustering, *168*
muscle form, 167
neuronal form, 168–169
opening and conformational change, 166–167
posttranslational modification, 167–168
structure, 163–164
ion selectivity and current flow determinants, 164–166
membrane-spanning segments, 164, *166*
Nieuwkoop's transformer, 293
Nissl substance, 42, 61, 66
Nitric oxide (NO), 858
guanylate cyclase regulation, 195–196
as neurotransmitter, *134–135*, 136
storage and release of, 134–135
synthesis, 134
termination action of, 135
Nitric oxide synthetase (NOS), 746
NMDA receptor, *see N*-Methyl-*D*-aspartate receptor
NMDA-receptor antagonis, *1032*
N-Methyl-*D*-aspartate (NMDA) receptor, 174–175, *174*
channel properties, 175
ligands, 174
long-term potentiation induction, 1019–1020
RNA splicing, 175
NO, *see* Nitric oxide
Nociception
axon reflex, 539
chemical sensitivity of nociceptors, 539
first and second pain, 538
human axis of pain, 546–547
nociceptors, 538
nonperceptual elements, 547–549
SI role, 546
spinal cord pathways, 545–546
ascending paths to thalamus, 546
spinothalamic neurons, 546
transient receptor potential receptors, 539
Nociceptor proteins, 539–540
Nodes of Ranvier, 64
Nonadrenergic noncholinergic (NANC) neurons, 746
Nonassociative learning, 1010
Nonmotor functions, basal ganglia in, 669–675
Nonrapid eye movement (NREM) sleep, 847–850, 854–855
Norepinephrine (NE)
biosynthesis, 120, *121*
neurochemical adaptation in addiction, 889
neuron, 851
release
autoreceptor regulation, 123–124
enzymatic inactivation, 124
neuronal transporters, 124–125
storage
vesicular monoamine transporters, 122–123
vesicular storage, 122
transporter, 124–125
NOS, *see* Nitric oxide synthetase
Notch, cellular determination, 313, 316, 322, 325–326
Notochord, evolution, 21, 23, *23–24*
NPY, *see* Neuropeptide Y
NREM sleep, *see* Nonrapid eye movement sleep
NSC, *see* Neural stem cell
NSF, *see N*-Ethylmaleimide-sensitive factor
NT, *see* Neurotensin
NTD, *see* Neural tube defect
NTFs, *see* Neurotrophic factors
Nucleus of the solitary tract (NST)
blood volume regulation, 791
satiety signaling, 770
taste processing, 518–519
visceral sensation, 720
Nucleus prepositus hypoglossi (NPH), *707*
Nystagmus, 688, 703

O

Obesity, 774
treatment of, *774–775*
Object recognition, 947
agnosia, 948–949
associative agnosia, 949–950, *949–950*
case study, *949*
inferior temporal cortex, 948, 950
development
dorsal visual stream and spatial attention, 924
face processing, 925–926
functional development, 923–924
ventral visual stream and object representation, 924
FFA, 961, *962*, *964*
human studies
dorsal and ventral processing streams, 959
LOC, 959
neuronal properties
experience effects on response properties, 958
face selective neurons, 956–957
invariant representations in occipitotemporal pathway, 954–956
occipitotemporal pathway cell sensitivity and face selectivity, 953–954
object knowledge storage, 965–966
pathway studies in nonhuman primates
classification of pathways, 950–953
interconnected cortical areas, 951–953, *952*
perception and recognition
category-selective regions, 963–964
category-specific representations, *963*, 965
prosopagnosia, 960–962
ventral pathway, 962–963
viewpoint dependence, *955–956*
Obsessive compulsive disorder (OCD), features of, *672–673*
Obstructive sleep apnea, *763*
Occipital cortex, 34
Occipitoparietal pathways, 950–951
Occipitotemporal pathways, 950–951
Ocular dominance columns, 448–449, *450*
Odor mapping, *525*
Odorant receptor (OR) gene, 522
OHCs, *see* Outer hair cells
Ohm's law, 99
OKR, *see* Optokinetic response
Olfaction, 520
adenylate cyclase signaling, 523
Drosophila system, 525–527
electro-olfactogram, 522
G protein-coupled receptors, 521–522, 528
hypothalamus role, 719
odor mapping, *525*
odor molecule classes, 520–521
olfactory bulb, 523
glomeruli activation and targeting by neurons, 524–527
olfactory cortex connections, 525
for smell, 524–525
synaptic process, 524
olfactory receptor neurons
Caenorhabditis elegans, *522*
gene expression, 522
projections to olfactory bulb, 523–524
transduction, 521–522

Olfaction (*Cont.*)
pheromone detection
accessory olfactory bulb, 527
accessory olfactory system comparison with main olfactory system, 529
humans, 528
vomeronasal organ
signal transduction, 528
structure, 527–528
TRPC2 knockout mouse, 528
Olfactory sensory neuron (OSN), 501, 503–504
Oligodendrocyte, myelin synthesis, 49–51
Olivary pretectal nucleus (OPN), 833
Olivocochlear (OC) neuron, 564–565
Ommatidia, 323, 578
Omnipause neuron (OPN), 705–706, *706*
On-off cells, 581
On-path dendritic inhibition, *238*, 239
OP-1, *see* Osteogenic protein-1
Opiates
neurobiological substrates, 885–886
peptide systems, 886
receptors, *885*
Optic ataxia, 959, 970–971, *971*
Optic disc, 583
Optic tectum, 707
Optic vesicle, 23
Optical disector, 931
Optical fractionator, neuron counting, *931*
Optogenetics, *213–214*
Optokinetic response (OKR), 697
gaze stabilization, 704–705
OR gene, *see* Odorant receptor gene
Orbitofrontal cortex, 656, 669, 673–674
Orexin, 851–852, *853*, 858
food intake regulation, 779
Orthodenticle genes, 298
Oscillators, 821
peripheral, 824
single-cell SCN, 821–824
Osmotic homeostasis, *see* Fluid intake
OSN, *see* Olfactory sensory neuron
Osteogenic protein-1 (OP-1), dendrite growth and branching regulation, 461
Otic rhombomere, 23–24, *24*
Otoacoustic emissions, *560*
Outer hair cells (OHCs), 557–559, *558–559*
OVLT, *see* Vascular organ of the lamina terminalis
Oxygen, chemoreceptor, 761–762
Oxytocin, 779, 814
salt appetite regulation, 795

P

P0, *see* Protein zero
Pacemaker, 821
in mammals, 820–821
Pacinian corpuscle, 534
PAG, *see* Periaqueductal gray
Pain, *see* Nociception
Paired helical filaments (PHF), 942
Paired pulse facilitation (PPF), 158, 227
Paired-pulse depression, 158
Papez's circuit, 720
Par proteins, 315
Parabionts, 773
Paradox, 667
Parafacial respiratory group (pFRG), 751, 754
Parahippocampal and retrosplenial cortex, 984
Parahippocampal place area (PPA), *962*, *964*, 984
Paramedian pontine reticular formation (PPRF), *702*, 705, *706–707*
Parasagittal stripes, *680*
Parasol cells, 584
Parasympathetic nervous system (PSNS), 735, *736*
autonomic innervation of, *735*
function, 734–735
postganglionic neurons, 738–739
Paraventricular hypothalamic nucleus (PVH), 720–721, *721*, 723, *725*, 726
Paraventricular nucleus (PVN), 745, 777, *778*, 813
Paravertebral ganglia, 731–732
Parietal cortex
anatomy, 34, 970
fronto-parietal attention control system, 994–998
lateral intraparietal area, 975–976
and visual attention, 710
spatial cognition decision processes, 977–978
injury effects, 970
constructional apraxia, 971–972, *972*
hemispatial neglect, 972–973
optic ataxia, 970–971, *971*
lesion studies, 974
neurons encode attributes, 974
spatial information eye movements, 976–977
ventral intraparietal area, 974–975
vision processing, 593–594
Parkinson's disease (PD), 1041
MPTP model, *667*
pathophysiology, *657*, 665–667
Pars compacta, substantia nigra
damage effects, 665–667, *666*
dopamine input to striatum, 660
spontaneous activity, 663
Pars reticulata, substantia nigra, 659–660, *709*, 711
damage effects, 665
spontaneous activity, 662–663
Parvocellular layer, lateral geniculate nucleus, 584, 586
Passive cable theory, 231
Patch-clamp, 215
channel opening and closing, 215–216
gating properties of ligand-gated channels, 217
IPSP, 220–222
multiple postsynaptic potential components, 222–223
nonlinear current-voltage relations, 220
null potential and slope of current-voltage relations, 217–218
recording, 244
statistical analysis, 216–217
summation of single-channel currents, 218–220
Path integration, 985
Pattern completion, 1039
Pattern separation, 1039
Pavlovian conditioning, *see* Classical conditioning
PCD, *see* Programmed cell death
PCP, *see* Planar cell polarity signal
PD, *see* Parkinson's disease
Pedunculated spines, *252*
Pelvic-hypogastric plexus, 733
Pentose phosphate pathway, 269–270
Peptide neurotransmitters
inactivation, 130–132
synthesis and storage, 130
Peptide receptors, 186
Peptide YY (PYY), 771
Perceptual illusion, 1097–1098
Perceptual learning, 1047–1048
Perforant path, 930
Perforated synapses, 43
Periaqueductal gray (PAG), somatosensory system, 547
Perikaryon, definition, 41
Period
circadian control, 825–826, 830
light-induced expression, 835–836
Peripheral nervous system (PNS), 72, 287
development, sensory neuron specification, 320–323
divisions, 32–33
neural crest, *see* Neural crest cells
neurogenesis, 347
Peripherin, 80
Perisylvian cortex, for language processing, 1062
Peroxisomal targeting signal 1 (PTS1), 74
Peroxisomes, mechanisms for, 73–74
Persistent vegetative state (PVS), *1094*
PET, *see* Positron emission tomography
PFC, *see* Prefrontal cortex
pFRG, *see* Parafacial respiratory group
Phase locking responses, for sound frequency, 563, *563*
Phenylethanolamine-*N*-methyltransferase (PNMT), catecholamine biosynthesis, 122
Pheromone detection
accessory olfactory bulb, 527
accessory olfactory system comparison with main olfactory system, 529
humans, 528
sexual behavior regulation, 816–817

vomeronasal organ
signal transduction, 528
structure, 527–528
TRPC2 knockout mouse, 528
Pheromones, definition, 529
PHF, *see* Paired helical filaments
Phosphatidylinositol 3-kinase (PI3K), neurotrophin receptor signaling, 423
Phospholipase C (PLC), G protein-coupled receptor signaling, 189, 193–195
Phosphorylation, *see* Protein phosphorylation
Phylogenetic development, 677–678
Pia, 35–36
Piccolo-bassoon transport vesicles (PTVs), 395
Picrotoxin, $GABA_A$ binding, 170
Pigment dispersing factor (PDF), 823
Pineal gland, 720–721, *722*
circadian regulation, 841
melatonin release, 841, 843
Pituitary, *see also specific hormones*
hypothalamic regulation, *see* Hypothalamus
neuroendocrine regulation, 723–724
PKA, *see* Protein kinase A
PKC, *see* Protein kinase C
Place cells, 985, 1036
Place learning, 1049
vs. response learning, *1042*
Placodes, 287
Planar cell polarity (PCP) signal, 295
Planum temporale, 915
Plasma osmolality, 772
Plateau potentials, 244, *625*
PLC, *see* Phospholipase C
PlexinA4 receptor, 400, *401*
Plexuses, 736
PLP, *see* Proteolipid protein
PME, *see* Prechordal mesendoderm
PNMT, *see* Phenylethanolamine-*N*-methyltransferase
PNS, *see* Peripheral nervous system
Pontine respiratory group (PRG), 751–752
Pop-out search, *1001*
Positron emission tomography (PET)
addiction studies, *892*
cerebral blood flow and energy metabolism, *266*
glucose uptake and metabolism in brain, 278–279
language studies, 1060
object recognition studies, 959, *960*, 961
sleep studies, 859
Postabsorptive state, 768, *768*
Posterior parietal cortex (PPC), 913
Posterior pituitary gland, 799
Postganglionic motor neurons, 730
Poststimulus time (PST) histogram, cochlear nucleus unit classification, *565*, 566–567
Postsynaptic cells, 441, 448, 453
Postsynaptic density (PSD), 175, 250, *252*, *395*
Postsynaptic neurotransmitter receptors, 399
Postsynaptic potentials (PSPs), 139, 158, 271, *see also specific potentials*
integration
spatial summation, 227, *227*
temporal summation, 226–227, *227*
metabotropic receptors and slow synaptic potential mediation, 224–226
patch-clamp, 215
channel opening and closing, 215–216
gating properties of ligand-gated channels, 217
IPSP, 220–222
multiple postsynaptic potential components, 222–223
nonlinear current-voltage relations, 220
null potential and slope of current-voltage relations, 217–218
statistical analysis, 216–217
summation of single-channel currents, 218–220
stretch reflex studies of ionotropic postsynaptic potentials, 211–214
Posttetanic potentiation (PTP), 158
Posttetanic potentiation (PTP)-3A, 398
Posture control, 605
medial postural system
balance and context-dependent postural strategies, 636
brain stem, 635
cervicocervico reflex, 635
vestibular and reticular nuclei, 631–632
vestibular apparatus, 632–633
vestibular damage and postural disorders, 636–637
vestibular plasticity, *637*
vestibular system neurons in, 633–634
vestibulocervical and vestibulospinal reflexes, 634–635
Potassium channel
action potential generation, 102–105
current voltage sensitivity and kinetics, 112–113
hyperpolarization-activated currents and rhythmic activity, 114–116
Potentiation, 158, *159*, 160
PP-1, *see* Protein phosphatase-1
PPA, *see* Parahippocampal place area
PPC, *see* Posterior parietal cortex
PPF, *see* Paired pulse facilitation
PPRF, *see* Paramedian pontine reticular formation
Prandial state, 768, *768*
PreBötzinger Complex (preBötC), 750, *753*, 753–754
Prechordal mesendoderm (PME), 293
Preclusters, *324*
Prefrontal cortex (PFC)
anatomy and organization, 1069–1071, *1070–1071*
diverse information to interact, 1070–1071
limbic system connectivity, 1070
motor system connectivity, 1070
attention deficit hyperactivity disorder defects, *1080–1081*
basal ganglia connectivity, 1070
development
dorsolateral prefrontal cortex, 922–923
visual search, 926
function theories
basal ganglia comparison, 1087
controlled processing flexibility and limited capacity, 1086–1087
multiple processing levels, 1085
rule representation, 1085–1086
somatic marker hypothesis, 1084–1085
top-down control, 1087–1088
working memory, 1083–1084
functional topography, *1076*
injury effects
humans
decision making, 1073–1074
disinhibition, 1072–1073
dysexecutive syndrome, 1071–1072
emotional impairment, 1073
planning impairment, 1073–1074
working memory impairment, 1074
nonhumam primates
delayed-response task, 1078
emotional impairment, 1077
rule-based tasks, 1076–1077
neuroimaging studies, 1074–1075
neurophysiology
multimodal properties, 1078
neuron activity
flexibility, 1083
reward encoding, 1083
sustaining, 1078–1080
task rules, 1080–1083
sensory system connectivity, 1069–1070
spatial cognition role, 981–982
Preganglionic neurons, 735–736
use acetylcholine, 738
Premotor cortex, spatial cognition role, 980
Prepyriform, *see* Pyriform
Presynaptic differentiation, 387
retrograde signaling and, 393
Prevertebral ganglia, 732, *732*
PRG, *see* Pontine respiratory group
Primates
communication systems in, 1055–1056
early, 909–913
evolution of, 906–909, *907*
prosimian, 906, *907*

Priming, of repetition, 152–153, 1048
Progesterone-response elements (PREs), 811
Programmed cell death (PCD), 405, *see also* Apoptosis
  development role in nervous system, 407–409
  evolution, 407
  functions in nervous system, 409–410
  mechanisms, 412–416
  modes, 410–412
  neuropathology and therapeutic targeting, 431–434
  neurotrophic hypothesis, 406
  quantification, 410
  regulation, 429–431
Projection interneuron, function, 26, *32*
Prolactin
  disruptions, 813
  release patterns, 840
Prolactin-inhibiting hormone, 800
Proliferative matrix cell movements, 351
Proneural clusters, 312
Proprioception
  complexity, 536
  Golgi tendon organ, 536
  muscle spindle, 536
Prosopagnosia, 960–961
Prostriate area, 904
Protein kinase A (PKA), *190*, 198–199
  cognition role, 201
  long-term sensitization in *Aplysia*, 1012–1014
  short-term sensitization in *Aplysia*, 1011
  signaling, 200
  sleep signaling, 859
  slow synaptic potential mediation, 224
Protein kinase C (PKC), *190*
  cognition role, 202
  phosphatidylinositol signaling, 200–201
  short-term sensitization in *Aplysia*, 1011
Protein kinase M (PKM), 1013
Protein phosphatase-1 (PP-1), structure and regulation, 202–203
Protein phosphorylation, 198
  cognitive kinases, 201–202
  kinase and phosphatase, 197–205
    networks of, 204
    spatial localization of, 201
  therapeutic targeting, 204
  tyrosine kinases, 202
Protein synthesis, 61–62, 65–75
  rough endoplasmic reticulum, 66–68
  trafficking, 68–70
Protein zero (P0), function, 51
Proteins, nociceptor, 539–540
Proteolipid protein (PLP), function, 51
PSD, *see* Postsynaptic density
Pseudostratified columnar epithelium, *348*
PSNS, *see* Parasympathetic nervous system
PSPs, *see* Postsynaptic potentials
Psychophysics, definition, 500
PTNs, *see* Pyramidal tract neurons
PTP, *see* Posttetanic potentiation
PTS1, *see* Peroxisomal targeting signal 1
PTVs, *see* Piccolo-bassoon transport vesicles
Publication, and knowledge advancement, 10–11
Purinergic receptors, 171, 184–185
Purkinje cells, 47, 231, 452, *453*, 610, 685–686
  in human, *232*
  mormyrid fish, *232*
Putamen neurons, 662
PVH, *see* Paraventricular hypothalamic nucleus
PVN, *see* Paraventricular nucleus
PVS, *see* Persistent vegetative state
Pyramidal cell, 47–48
Pyramidal neuron, 249, *250*
Pyramidal tract, 639, *640*
Pyramidal tract neurons (PTNs), 646
Pyriform, 525
PYY, *see* Peptide YY

Q

Quail cells, *341*
Quantal analysis, 155
  Katz model
    exceptions, 157
    standard, 157–158
  standard model of transmission, 155–157
Quantal release, 139–142
Quantal transmission, 155–157, *156*

R

RA, *see* Rapidly adapting; Retinoic acid
Rab3, *147*, 148–149, 152
Rabphilin, 148, *149*
Radial glial fibers, neuronal migration
  cerebellar system, 357
  cortical interneurons, 356
  layer 1 cells, *349*
  lineage analysis, 355–356
  RMS, 356–357
Rall's theory, 231
Raphe interpositus (RIP), *706*
Rapid eye movement (REM) sleep, 847–850, 855–858, *857*
Rapid eye movement sleep behavior disorder (RBD), 863
Rapidly adapting (RA), 533
Rapidly adapting pulmonary stretch receptors (RARs), 757
Rapsyn, 385, *386–387*, 392–393
  mutant mice, 392
  postsynaptic differentiation role, 392–393
RAS, *see* Reticular activating system
rCBF, *see* Regional cerebral blood flow
Reactive oxygen species (ROS), generation and glutathione protection, 270–271
Rebound firing, 686
Receptor desensitization, 158–160
Receptor potentials
  basolateral membrane shape, 556–557
  evoked in stereocilia, 555
Receptor protein tyrosine phosphatases (RPTPs), 397
Recognition memory signals, in monkey hippocampus, *1037*
Recognition memory tasks, *1034*
Recurrent inhibition, 626
Refractory periods, 105–106
Regional cerebral blood flow (rCBF), 1060
Regional identity genes, *314*
REM sleep, *see* Rapid eye movement sleep
Renshaw cell, 615, 626
Repetition priming, 1048
Reproduction, hypothalamic control of, 808–812
RER, *see* Rough endoplasmic reticulum
Respiratory dysfunction, in sleep, 862–863
Respiratory rhythm generation, 753–756
Response learning, *1042*, 1049
  place learning *vs.*, *1042*
Restless leg syndrome (RLS), 864
Reticular activating system (RAS), consciousness, 1095
Reticular formation
  reticulospinal, 631, 639
  vestibulospinal, 631, 639
Retina, 580–581
  amacrine cell, 582–583
  bipolar cell, 581
  cellular determination
    development, *327*
    Drosophila, 323
    Hh-mediated feedback mechanism, 325
    ommatidia, 323–324, *324*
    R7 cell, 325
    vertebrate, *327*, 328
    Xenopus and zebra fish, *326*
  cones, 581
  horizontal cell, 582–583
  retinogeniculocortical pathway
    extrastriate visual areas, 592–593
    lateral geniculate nucleus, 586–587
    parietal cortex, 593–594
    primary visual cortex
      cat, 587–589
      macaque, 589
    visual cortex
      columnar structure, 589–590
      cytochrome oxidase staining, 592
      optical imaging studies, 590–592
  rods, 581
  suprachiasmatic nucleus entrainment
    input to nucleus, 835
    photoreceptive system, 832–835
Retinal ganglion cells (RGCs), 380, 583
  in cat, *232*, 583–584
  classification, 583

mapping by Kuffer, *578*
projections in primates, 584
Retinoic acid (RA), 293–294
Retrograde amnesia, 1039–1040
Retrograde organizers, 396
Retrograde signals, 396–397, *397*
Retrosplenial cortex (RSC), 984
Rett syndrome, *764*
Reversal potential, 218
Reward, 871–879
and addiction, *see* Addiction
catecholamine hypothesis of rewarding brain stimulation, 877
dopaminergic neurons, 876
intracranial stimulation studies, 876, *878*
motivated behavior, 874–876
psychophysics of hedonism, 876, *878*
RGCs, *see* Retinal ganglion cells
Rho, dendritic cytoskeleton structure control, *464–465*
Rhodopsin, 196
Rhombic lip (RL), 156
Rhombomeres, 310
RIP, *see* Raphe interpositus
RNA polymerase II, 205
Robo, dendrite growth and branching regulation, 462
Rod, eye, 581
ROS, *see* Reactive oxygen species
Rostral interstitial nucleus, 705
Rostral interstitial nucleus of medial longitudinal fasciculitis (riMLF), *702*, 705
Rostral migratory stream (RMS), 158, 356–357, *356*
Rostral ventral respiratory group (rVRG), 751
Rostral ventrolateral medulla (RVLM), 759, 761
Rostrocaudal axis, *21*
Rough endoplasmic reticulum (RER)
protein synthesis, 66–68
protein trafficking, 68–70
RPTPs, *see* Receptor protein tyrosine phosphatases
RSC, *see* Retrosplenial cortex

S

S-100, astrocyte function, 51
Saccades, 705
Sacculus, vestibular apparatus, 632
Salt appetite, 783, 792
adrenocortical dysfunction, *794*
aldosterone secretion and, 793–795
atrial natriuretic peptide, 795
hypovolemia, *796*
oxytocin, 795
Saltatory conduction, 49
SAR, *see* Slowly adapting pulmonary stretch receptor
Sarcomeres, 616
SBMA, *see* Spinal and bulbar muscular atrophy
SC, *see* Superior colliculus
SCA, *see* Spinocerebellar ataxias
Scaffolding molecules, *396–397*, 398
Schizophrenia, 938
dopamine hypothesis, 938
genetics, 939–940
neurodevelopmental hypothesis, 939
neuropathology, 938–939
Schwann cells, 343, 387, *389*
myelin synthesis, 49–51
Sclerotome, *345*
SCN, *see* Suprachiasmatic nucleus
SCO, *see* Subcommissural organ
SEF, *see* Supplementary eye field
Segmental migration
Eph receptors, *346*
ganglia, 346
neural crest migration, 346–347
neurogenesis in PNS, 347
segmentation, 345
somites, *345*
Self-avoidance, organization of dendritic fields, 469
Semaphorin, growth cone guidance, 370–372
Semicircular canals, vestibular apparatus, 632
Sensation, 499
Sensilla, 322
Sensitive period, plasticity after, 486
Sensitization, 1010
long-term, 1012–1014
short-term, 1011–1012
temporal domains for memory of, 1014
Sensorimotor disturbances, 864
Sensorimotor neurons, 850
Sensorin, synthesis of, 1013
Sensorineural hearing loss, 559
Sensory cortex, *see also specific areas*
association areas, 507
columnar organization, 508–509
neuron response complexity, 510
prefrontal cortex connectivity, 1069–1070
primary areas, 507
response mappings and plasticity, 508
stereotyped connections, 509–510
Sensory neurons (SN), 1011
nerve net, 16–17, *17*
Sensory organ progenitor (SOP), 322
Sensory placodes, 320
Sensory systems, 499–500, *see also specific systems*
central pathways and processing
axons, 505–506
multiple maps and parallel pathways, 507
thalamic nuclei, 506–507
convergence and divergence sites, 502–503
embryonic origin, 503
input range specificity, 501
peripheral organization and processing
labeled line principle, 503–504
lateral mechanisms for sensitivity to contrast, 505
neural signaling codes, 505
topographic projections, 504
position patterns and density, 502
psychophysics, 500
types, by system, 501
Serotonin, 125–126
5-HT3 properties, 169
neurochemical adaptation in addiction, 888–889
receptors, 185
short-term sensitization in *Aplysia*, 1011–1012
Sevenless mutants, 325, *325*
Sex differences, neurodevelopmental disorders, 940
Sexual behavior
hormonal regulation, 815–816
pheromone regulation, 816–817
Sexual differentiation, of brain, 815
SFO, *see* Subfornical organ
Sharp wave ripples (SWRs), 1040
Short-term potentiation, breathing, 762
Short-term sensitization, 1011–1012
Shy–Drager syndrome, *744*
Sidekicks, 400
SIDS, *see* Sudden infant death syndrome
Sight, *see* Vision
Signal transducers and activators of transcription (STATs), 206
Signal transduction, 892–896
receptor cells, taste, 516
in VNO, 528
Silent synapses, 157
Simultaneous somatic current injection, *250*
Single unit recording studies, of fronto-parietal attentional control system, 994–998
Single-source divergent neurons, 6
Size principle, of motor neurons, 619–620
Skeletal motor system, 729
Sleep, 847–849, 865–867
brain stem regulatory system, REM sleep initiation, 855–858
central autonomic control systems, 849
changes over human lifespan, 861–862
circadian disorders, *829–830*
disorders
behavior state control disruption, 862
motor disturbances, 863–864
respiratory dysfunction, 862–863
dreaming, 860–861
electroencephalography, 847
electromagnetic physiology, 859
memory interdependence, 865–866, *865*
neuroimaging of, 859–860

Sleep (*Cont.*)
neurotransmitter systems in behavioral state control, 858–859
onset regulation, 852–854
phases
nonrapid eye movement sleep, 847–850, 854–855
REM sleep, 847–850, 855–858, *857*
phylogeny of, 862
sleep–memory interdependence, 865–866
suprachiasmatic nucleus and sleep-wake cycle regulation, 838–839
thalamo-cortical interactions in nonrapid eye movement sleep rhythms, 854–855
VLPO control of sleep-wake cycles, 852
wakening maintenance, 851–852
Sleep apnea, *763*, 862, *864*
Slit
dendrite growth and branching regulation, 462
growth cone guidance, 372
Slowly adapting afferents
receptor density, 536
type I, 535
type II, 535–536
Slowly adapting pulmonary stretch receptors (SARs), 756–757
SMA, *see* Spinal muscular atrophy; Supplementary motor area
Small aspiny neuron, striatum, 655
SMN, *see* Survival of motor neuron
Smooth endoplasmic reticulum (SER), *252*
SNS, *see* Sympathetic nervous system
Sodium channel, 107–108
action potential generation, 102–105
current transience and persistence, 112
hyperpolarization-activated currents and rhythmic activity, 114–116
salt taste detection, 516
Sodium-potassium ATPase, *95*
Solar plexus, 732
Solitary tract nucleus, *see* Nucleus of the solitary tract
Soluble NSF attachment proteins (SNAPs), 69
membrane fusion, 149–152, *150*
protein trafficking, 69
Soluble NSF attachment proteins receptors (SNAREs), 69
membrane fusion, 149–152, *150*
protein trafficking, 69
Soma, *see* Perikaryon
Somatic membrane potential, *233*
Somatic motor pattern, *730*, *741*
Somatic recordings, *240*, *245*
Somatosensory, 905, *905*, 912, *912*
Somatosensory cortex
central nervous system components
ascending paths, 540–542
lateral cervical nucleus, 544
mechanosensory paths, 542–544
thalamic segregation of place and modality, 544–545
cortical representation of touch
SI lateral inhibition, 549–550
columnar organization, 549–550
orientation selectivity, 549
serial processing, 549
SII
attention-modulated responses, 550
body representations, 550
receptive fields, 550
nociception
axon reflex, 539
chemical sensitivity of nociceptors, 539
first and second pain, 538
human axis of pain, 546–547
nociceptors, 538
nonperceptual elements, 547–549
SI role, 546
spinal cord pathways, 545–546
ascending paths to thalamus, 546
spinothalamic neurons, 546
transient receptor potential receptors, 539
peripheral mechanisms
afferent fibers and roles, *533*
hair follicle afferents, 536
mechanosensory axon, types in, 533–536
Meissner corpuscle, 534–535
Merkel's disk, 535
Pacinian corpuscle, 534
receptors and ganglion cells, 531–532
slowly adapting afferents
receptor density, 536
type I, 535
type II, 535–536
proprioception
complexity, 536
Golgi tendon organ, 536
muscle spindle, 536
SI, 540
SII, 540
touch
SI lateral inhibition, 549–550
columnar organization, 549–550
orientation selectivity, 549
serial processing, 549
SII
attention-modulated responses, 550
body representations, 550
receptive fields, 550
trigeminal system, *548–549*
Somatosensory system, labeled lines principle, 503–504, *504*
Somatostatin, 806–807
disruptions, 808
growth hormone-releasing hormone, 807–808
Somatotopic organization, 642–645
Song learning, and memory, 482–483
SOP, *see* Sensory organ progenitor
Sound frequency, 563
Sound localization
binaural cues to determining, 569–572, *569*
cortical fields and, 573
of owl, 483–487
plasticity mechanisms, 486
sensitive period for neuronal adjustments, 485–486
Spatial buffering, astrocyte, 52
Spatial cognition
frontal cortex, functional overview, 978–979
hippocampus role
declarative spatial memory, 984
memory formation, 984
neuron place fields, 984–985
parahippocampal and retrosplenial cortex, 984
motor action, 986
motor cortex, 979–980
neural systems
dorsal stream areas, 969
spatial processing streams, 969–970
parietal cortex, 970
decision processes, 977–978
lateral intraparietal area, 975–976
prefrontal cortex role, 981–982
premotor cortex role, 980
supplementary eye field role, 980–981
Spatial neglect, 990
Spemann's organizer, 287–288, 293
Spinal and bulbar muscular atrophy (SBMA), *622*
Spinal cord, 33
dorsal, *330*
injury, *627*
interneuron cell, 329
motor neurons, 328
pain pathways, 545–546
ascending paths to thalamus, 546
spinothalamic neurons, 546
segregation of place and modality in pain, 544–545
ventral neural tube progenitors, 328, *329*
vertebrate, *329*
Spinal interneurons, 622
central pattern generator network models, 622–624
excitatory interneurons, 624–625
left–right coordination, 626
reciprocal inhibition between flexor and extensor muscles, 625–626
Renshaw inhibition, 626
Spinal muscular atrophy (SMA), *622*
Spinal nerves, AP axis of, 301–302
Spine apparatus, 44, 251
Spineless, dendrite development role in *Drosophila*, 459
Spinocerebellar ataxias (SCA), *937*

Spiny motor neuron, 48
Spiny stellate cells, 47
SPL, *see* Superior parietal lobule
Standard Katz model, 157–158
Stapedectomy, 553
STATs, *see* Signal transducers and activators of transcription
Stellate ganglion, 732
Stem amniotes, 901
Stereocilia, 502
  linking transducer gating, 555–556, *557*
  receptor potentials evoked in, 555
Steroid hormone receptors, transcription regulation, 208
STG, *see* Superior temporal gyrus
STN, *see* Subthalamic nucleus
Stress, hypothalamic control of, 803–804
Stretch reflex, *601*
Striatal neurons, 661–662
Striatum, *see* Basal ganglia
Striatum, memory system, 1041, *1041*
  for habits, actions, and outcomes, 1042–1043
  identification, 1041–1042
Striosomes, 658
Stroop task, *1073*
Subcommissural organ (SCO), 721, *722*
Subfornical organ (SFO), 721, *722*, 725–726, *789*, 792
Sublaminae, 400
Substance dependence
  dopamine and serotonin systems, 888
  neuroadaptive changes, 888–889
  signal transduction mechanisms, 892–896
Substantia nigra, *see* Basal ganglia
Substantia nigra (SN)
  globus pallidus, internal segment, 659–660, *659*
  pars compacta (SNpc)
    damage effects, 665–667, *666*
    dopamine input to striatum, 660
    spontaneous activity, 663
  pars reticula (SNpr)
    damage effects, 665
    spontaneous activity, 662–663
Substrate phase, 768, *769*
Subthalamic nucleus (STN)
  damage effects, 664–665
  inputs from frontal lobe, 658
  role of, *674–675*
  spontaneous activity, 662
β Subunit of integrin, 340
Sudden infant death syndrome (SIDS), *763*
Superior cervical ganglion, 732
  autonomic innervation of, *735*
Superior colliculus (SC), gaze shift control, 707–711, *709*
Superior mesenteric ganglion, 733
Superior parietal lobule (SPL), attention and control system, 992
Superior temporal gyrus (STG), 1062
Superior temporal sulcus, *957*
Supplementary eye field (SEF), *709*, 710
  attention control system, 992
  spatial cognition role, 980–981
Supplementary motor area (SMA), prefrontal cortex connectivity, 1069
Suppressive masking, 564
Suprachiasmatic nucleus (SCN)
  entrainment, 832, 835–836
  neuroanatomy, 819
  oscillator and pacemaker function, 821
  outputs
    amplification and divergence, 837–838
    autonomic function regulation, 840–841
    humoral outputs, 837
    neuroendocrine function regulation, 839–840
    pineal function regulation, 837
    sleep-wake cycle regulation, 838–839
  retina
    input to nucleus, 835
    photoreceptive system, 832–835
  rhythmic firing of neurons, 836–837
  single-cell oscillators, 821–824
  visual projections, 720, 723
Supraoptic nucleus (SON), 813
Survival of motor neuron (SMN), *622*
SV, *see* Synaptic vesicles
SWRs, *see* Sharp wave ripples
Sympathetic ganglia, 343
Sympathetic nervous system (SNS)
  fight-or-flight response, 730
  postganglionic neurons, 738–739
  preganglionic neurons, 731–733
Sympathetic premotor neurons, 759–760
Synapse, 4, 43
  components, 43
  development
    changes, 921
    glial-derived signals for, 400
  elimination, 437–443
  formation
    and agrin-mediated signaling for MuSK, 390–391
    in central nervous system, 394–401
  functional categories, 4–5
  neurotransmitter release, *see* Neurotransmitters
  structural categories, 4
Synapse-defective-2, 398
Synapse-organizing signals, at CNS synapses, *397*
Synaptic cleft, 140, 154–155
Synaptic connectivity, 394
Synaptic depression, 156, 158–160, *159–160*
Synaptic differentiation, for synaptic basal lamina signals, 387–389
Synaptic integration
  in active dendrites
    dendritic spikes, 244–249
    inhibition and excitability, 249–250
    intracellular recordings, 239
    patch-clamp recording, 239
    voltage-gated channels, 242–244
    weakly excitable dendrites, 240–242
  in passive dendrites
    filtering functional implications, 239
    interactions between synapses, 237–239
    potentials at soma, 234–237
    synaptic kinetics on dendritic filtering, 237
Synaptic null potential, 218
Synaptic organization, 396
  morphogens retrograde signals for, 396–397
Synaptic plasticity, 227–229
  neurotrophin role, 422–423
Synaptic specificity
  model mechanisms for, *401*
  in surface recognition molecules, 400–401
Synaptic transmission, *243*
Synaptic vesicles (SV), 386, *396*
  exocytosis of, *145*
  fusion of, *151*
  life cycle of, *142*
  neurotransmitters, 145
  proteins, 147–149
  structure and topology, *148*
  VAMP, 150
Synaptobrevin, *69*, 70, *150*
  membrane fusion, 150
  protein trafficking, 70
Synaptophysin, 147
Syntactic operations, 1064–1066
Syrinx, 1056
Systems theory, *582*

T

Taste, 513
  central transmission, 518–519
  encoding
    central nervous system, 520
    labeled lines, 519
    periphery coding, 519–520
    population coding, 519
  gustatory afferent neurons, *518*, 519
  receptor cells
    acid depolarization, 516–517
    G protein-coupled receptors, 514, 516
    neuron activation and taste behavior generation, 517–518
    receptor gene organization, 517
    salt taste detection, 516
    signal transduction, 516
    T1R receptor family, 515
    T2R receptor family, 515–516
    tuning in *Drosophila*, *518*

Taste (*Cont.*)
taste buds
clusters, 514
innervation, 514
TCA cycle, *see* Tricarboxylic acid cycle
TEA potassium, *see* Tetraethylammonium
Temperament
plasticity mechanisms, 490–491
sensitive period for shaping, 489–490
Temporal cortex, 34
Temporal lobes, 938, *952*, 953, 984–986, 1031
Testosterone, 811
Tetraethylammonium (TEA), potassium, current inhibition, 103
Δ^9-Tetrahydrocannabinol (THC), 886, *894*
Tetrodotoxin (TTX), sodium current inhibition, 103
TGN, *see trans*-Golgi network
TH, *see* Tyrosine hydroxylase
Thalamus, *see also specific nuclei*
gaze movements control, 711
sensory nuclei, 506–507
sleep thalamo-cortical interactions in nonrapid eye movement sleep rhythms, 854–855
spinal cord pathways in pain, 545–546
ascending paths to thalamus, 546
spinothalamic neurons, 546
synapse elimination, 451–452
Thirst appetite
AVP secretion, 787–789, 792–793
hypovolemia, 792–793, *796*
osmotic dilution, 792–793
osmotic homeostasis regulation, 785
Thoracolumbar division, 731
Thrombospondin (TSP), 400
Thyroid hormone, 805–806
receptors, 806
Thyroid-stimulating hormone (TSH), release patterns, 805–806
Thyrotropin-releasing hormone (TRH), release patterns, 805–806
Tiling, 373, 467–468
Titubation, definition, *688*
TMS, *see* Transcranial magnetic stimulation
Tonically active neurons (TANs), 662
Topographic map, of "retinotectal" projection, 380–382
Topographic pattern, of neurons, 504
Torpedo electric organ, 385, *386*, 389–390
Touch dome, 535
Tourette syndrome (TS), features of, *673–674*
Tower of London task, *1073*
Transcranial magnetic stimulation (TMS), *627*
Transcription
AP-1, 206–207
cAMP and Ca^{2+} in, 206
cytokine induction, 207–208
factors in long-term memory, 1012
regulation by signaling, 206
signal-regulated, 205–206
Transcription, circadian feedback loop, *829*
Transducin, 581
Transforming growth factor β (TGF-β), 1012–1013
trans-Golgi network (TGN), 70–72
Transient receptor potential (TRP) receptors, nociceptors, 539
Translational cytoplasm, 61
Translocating chain-associating membrane protein (TRAM), *67*
Translocon-associated protein (TRAP), *67*
Treadmilling, 374
TRH, *see* Thyrotropin-releasing hormone
Tricarboxylic acid cycle (TCA), *263*, 269
Trigeminal system, somatosensory system, *548–549*
Triglyceride, 768
Tripartite junctions, *395*
Triplet repeat diseases, *936–937*
TrkC receptor, 397
Trophic factors, and programmed cell death, 431–434
TRP, *see* Transient receptor potential
Trunk neural crest cells, *345*
TS, *see* Tourette syndrome
TSH, *see* Thyroid-stimulating hormone
TTX, *see* Tetrodotoxin
Tuberomammilary nucleus (TMN), 851
Tubulin, posttranslational modification, *77*
Tuning curve, *505*
Two-photon glutamate uncaging, *243*
Two-tone suppression, 564
Tyrosine hydroxylase (TH), catecholamine biosynthesis, 120–122

U

Ultradian patterns, neurohormone release, 802–803
Unconditioned response (UR), 1010
Unconditioned stimulus (US), 1009, *1010*
Unconventional transmitters, 132–136
Unipolar brush cell, *682*, 683
Utricle, vestibular apparatus, 632
Utrophin, 392–393

V

VAChT, *see* Vesicular cholinergic transporter
Vagal neural crest cells, *344*
Vagotomy, 770
Vascular endothelial growth factor A (VEGFA), 373
Vascular organ of the lamina terminalis (OVLT), 721, *722*, 725–726
osmoreceptor cells, 787–789
Vasoactive intestinal peptide (VIP), 821, *838*
Vasoactive intestinal polypeptide (VIP), and sleep control, 858
Vasopressin, *see* Arginine vasopressin
Venous sinus, 37
Ventilation, experimental analysis of, *751*
Ventral caudal, somatosensory system, 540
Ventral intraparietal area (VIP), and spatial cognition, 974–975
Ventral organizers, DV axis, 302–303
Ventral pathway, 1063
Ventral processing streams, 959, *960*
Ventral respiratory column (VRC), 751, *752*
Ventral respiratory group (VRG), 751
Ventral system, of basal ganglia, *670*
Ventral tegmental area (VTA), *670*
Ventrolateral hypothalamus (VLH), 776
Ventrolateral preoptic area (VLPO), control of sleep-wake cycles, 852
Ventromedial hypothalamus (VMH), *773*, 776
Ventroposteromedial nucleus of the thalamus (VPMpc), taste processing, 519
Vergence eye movement, 699
Versional eye movement, 699
Vertebral artery, 36
Vertebrate hindbrain, 296
Vesicle docking, 152–153
Vesicular cholinergic transporter (VAChT), 127
Vesicular monoamine transporter (VMAT), 122–123
Vestibular canals, and otoliths, *633*
Vestibular nuclei (VN), 689, *707*
Vestibular system
apparatus, 632–633
balance and context-dependent postural strategies, 636
damage and postural disorders, 636–637
neurons in, 633–634
plasticity, *637*
vestibular and reticular nuclei, 631–633
vestibulocervical and vestibulospinal reflexes, 634–635
Vestibulo-ocular reflex (VOR), 697
gaze stabilization, 703–704
VIP, *see* Vasoactive intestinal peptide; Vasoactive intestinal polypeptide; Ventral intraparietal area
Visceral afferents, 740–743
Visceral motor pattern, *730*, *732*
Vision, 577, *see also* Eye, Object recognition, Spatial cognition
color vision defects, *584*
color-selective response in P pathway, 585–586
development of visual system, 923–924
dorsal visual stream and spatial attention, 924
ventral visual stream and object representation, 924
eye, 579–580
M cells and contrast sensitivity, 586
photoreceptors, 581
receptive fields, 577

response properties of P an M pathways, 584–585
retina structure and cells, 580–581
retinal ganglion cell, *see* Retinal ganglion cell
retinogeniculocortical pathway
extrastriate visual areas, 592–593
lateral geniculate nucleus structure, 586–587
parietal cortex, 593–594
primary visual cortex
cat, 587–589
macaque, 589
visual cortex
columnar structure, 589–590
cytochrome oxidase staining, 592
optical imaging studies, 590–592
visuomotor coordination and reaction time, 608–611
Visual agnosia, 948–949
associative, *948*, 949–950, *950*
invariant object representations, 949–950
Visual cortex
experience effects on functional organization, 487
critical period, 487–489
plasticity mechanisms, 488–489
neural activation, 992–994, *992–993*
synapse elimination, 448–451
Visual masking, effect of, *1100*
Visual search paradigm, 1001
Visuospatial cognition, 970, 974, *978*
VLH, *see* Ventrolateral hypothalamus
VLPO, *see* Ventrolateral preoptic area
VMAT, *see* Vesicular monoamine transporter
VMH, *see* Ventromedial hypothalamus
VN, *see* Vestibular nuclei
VNO, *see* Vomeronasal organ
Vocal learning, *481*, 482–483
Voltage-clamp, principles, *103*
Voltage-gated calcium channels (VGCCs), *397*, 1018, *1019*, 1020
Voltage-gated channels
immunostaining, 242
nonuniform expression, 244
Volume homeostasis, *see* Fluid intake
Vomeronasal organ (VNO), 527–528
signal transduction, 528
TRPC2, 528
VOR, *see* Vestibulo-ocular reflex
VPMpc, *see* Ventroposteromedial nucleus of the thalamus
VRC, *see* Ventral respiratory column

W

Waking, and reticular formation, 851–852
Warts, dendrite development role in *Drosophila*, 460
WASP, *see* Wiskott–Aldrich syndrome protein
Water channels, *786*
Water intake, *see* Fluid intake
Water intoxication, *791*
Weakly excitable dendrites
backpropagating action potentials, *241*, 242
dendritic recordings, 241, *241*
soma and, 240
Weather prediction task, *1042*
Werdnig–Hoffman disease, *622*
Wernicke's area, *1060*
White matter
developmental changes, 922
spinal cord, *26*, 33
Wisconsin card sorting test (WCST), 939, 1072, 1074–1077
Wiskott–Aldrich syndrome protein (WASP), 352
Wnt
dendrite growth and branching regulation, 462
growth cone guidance, 372
Word recognition, 1062–1064
Working memory, 1074, 1083–1084

X

X cells, *582*

This page intentionally left blank